THREE ESSAYS ON THE ECONOMICS OF EDUCATION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Leigh Ramsay Wedenoja

August 2017

THREE ESSAYS ON THE ECONOMICS OF EDUCATION

Leigh Ramsay Wedenoja, Ph.D.

Cornell University 2017

Despite large measured returns to completing a high school diploma, the US has a dropout rate among the highest in the industrialized world and, while there are many studies of dropout, we still have limited knowledge of the process through which students dropout and the causal evidence of effective dropout prevention strategies is sparse. This dissertation studies the decisions students make while in high school, and how those decisions are related to dropout, academic achievement, and health. Specifically, it focuses on the relationship between high school attendance behavior, academic achievement, and dropout and the relationship between health care provision, teen pregnancy, and dropout.

Chapters 1 and 2 focus on how attendance and truancy affect students' ability to graduate from high school. I use a novel dataset of daily level student attendance data for multiple cohorts of ninth graders in a Large Urban School District (LUSD) to follow individual students' attendance decisions through their high school career. Chapter 1 provides the first detailed description of the daily patterns of high school attendance. The chapter examines how measures of attendance intensity, aggregate attendance, and attendance type evolve over the course of the school year and their relationship to student achievement. I find that, even given a total number of absences, the number of truant absences is the most predictive of negative educational outcomes. Truant absences become more frequent over the course of the school year, more persistent, and students

spend longer spells outside of school. These patterns are more intense for students in lower performing high schools and who have lower grades and test scores. Truancy predicts dropout and low test scores as early as the first month of high school and can serve as a valuable early warning indicator for school administrators. These patterns also provide suggestive evidence that attendance is an input into the education production function and insight into the dynamics of high school disengagement before students dropout.

In Chapter 2 I model dropout as the outcome of a series of small investment decisions, specifically the decision of a student to attend school or be truant on a given day, rather than dropout as a one-time decision based on the patterns of attendance detailed in Chapter 1. Students' daily attendance decisions not only affect their ability to graduate, but also the costs they face on future days and the quality of the diploma they can earn. I find that, once students are subject to the plausibly exogenous timing of an opportunity cost increase, they are truant more frequently and the accumulation of truancies occurs at a faster rate. A student who barely misses the school cut off age is 12% more likely to be truant in 12th grade, and a student with existing attendance problems is up to 69% more likely to be truant after only 3 months of increased schooling cost. Students with higher levels of truancy each year, especially truancies late in the school year, are less likely to return to school the following fall and if they do return are more likely to transfer into a non-traditional diploma program. These results help explain the persistence of the US's high dropout rate and suggest that policies to reduce the daily attendance costs faced by students, even late in high school, may be most effective at dropout prevention.

Chapter 3 turns its focus away from high school attendance to high school students' access to health care. I, along with coauthors Michael Lovenheim and

Randall Reback, explore whether teenagers' access to primary health care influences their fertility and educational attainment. We study how the significant expansion of school-based health centers (SBHCs) in the United States since the early 1990's has affected teen birth and school dropout rates. Our results indicate that school-based health centers have a negative effect on teen birth rates: adding services equivalent to the average SBHC reduces the 15-18 year old birth rate by 5%. The effects are largest among younger teens and among African Americans and Hispanics. However, primary care health services do not reduce dropout rates by very much despite the sizable reductions in teen birth rates.

BIOGRAPHICAL SKETCH

Leigh Wedenoja holds a Bachelors degree in Economics and Latin American & Caribbean Studies from the University of Michigan and a Certificate in Contemporary Studies from La Pontificia Universidad Católica de Chile. Prior to entering the Economics Ph.D. program at Cornell, Leigh spent a year in AmeriCorps and served with City Year Detroit. It was working with middle school students in the Detroit Public Schools that sparked Leigh's interest in the economics of education and inspired her to pursue research into how adolescents make decisions about their education, and why so many talented students fail to complete high school. Following graduation, Leigh will continue her education research by spending two years as a Postdoctoral Research Associate at Brown University in the Department of Education.

This document is dedicated to my parents and sisters, without whose love and support it would not exist, and especially to my mother who wrote her dissertation on the same desk I wrote mine.

ACKNOWLEDGEMENTS

I am deeply grateful to my committee, Michael Lovenheim, Ted O'Donoghue, and Kevin Hallock, for all of the time, effort, and energy they have spent on me throughout my career at Cornell. When I presented plans for this dissertation to my committee, they told me that it was a risky and ambitious project that could have any number of things go wrong. Either despite of, or because of that, they encouraged this project. When things inevitably did go wrong and research became difficult, it was their support and guidance that got me through. I am especially grateful to Mike, who agreed to chair my committee despite having never chaired one before, and has always pushed me to be a better researcher, to think critically, and to use commas correctly when I write.

I also benefited immensely from discussions with my fellow Ph.D. students, especially the participants in the Behavioral Economics Research Group, The Works in Progress Seminar, the Labor Works in Progress Seminar, and the Labor Economics and Policy Analysis and Management Seminars. I could not have asked for better people to spend grad school with.

This project would not have been possible without the support of a Large Urban School District and the district staff who lobbied the school board on my behalf. Nor would it have been possible without the staff of The Cornell Restricted Access Data Center, who were willing to jump through hoops to host my data and facilitate my research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

HIGH SCHOOL ATTENDANCE PATTERNS AND STUDENT OUTCOMES

1.1 Introduction

The measured benefits of graduating high school compared to dropping out or earning a GED are immense. Students who stay in school longer have higher incomes, better health, and are less likely to be incarcerated or unemployed (Angrist & Krueger, 1991; Cameron & Heckman, 1993; Card, 2001; Oreopoulos, 2007; Lochner & Moretti, 2004; Muenning, 2008.) However, despite these benefits, the graduation rate in the US remains one of the lowest in the developing world with estimates varying from 20% to 30% of all students. One of the strongest predictors of high school graduation is school attendance. There is a well documented relationship between student attendance and academic outcomes beginning as early as elementary and middle school. Students who are in school more often perform better on tests, are more likely to graduate, and are more likely to attend college. Schools with higher rates of average attendance also perform better on other school accountability measures (Barrington and Hendricks, 1989; Lamdin, 1996; Peterson and Colangelo, 1996; Strickland, 1998; Stanca, 2006; Dobkin et al., 2010; Douglas, 2004; Marbyurger, 2006; Neil et al, 2008).

Education policy and research has focused primarily on two aggregate measures of attendance. The first is the averaged yearly attendance rate, which is the average percentage of registered students present each day. This number is over 90% for the vast majority of schools (Attendance Works, 2015; Balfanz & Byrnes, 2012). More recently the focus has expanded to the rate of chronic absenteeism. Chronic absenteeism is defined as the percentage of enrolled students at a school who miss more than a certain cutoff of school days for either excused or unexcused absences.

While it is well established that higher aggregate attendance is correlated with achievement, how attendance fits into the education production function and the mechanisms through which attendance affects student achievement are still open questions. This paper expands the attendance literature by providing the first detailed description of the daily level attendance patterns of high school students and the correlation between those patterns and school completion and student achievement. Unlike previous studies of student attendance which are limited by aggregate measures of attendance, I employ a novel detailed administrative dataset of daily level student attendance data for a Large Urban School District (LUSD). The data provide detailed information on individual students including whether they were in school each day and, if not, the reason they were absent. The data contain the full high school attendance record for over 38,000 high school students and follows 5 cohorts of ninth graders until they graduate or leave the school district.

The relationship of different types of attendance patterns and measures to education outcomes provides insight into the role attendance plays in the education production function and student education investment decisions. Patterns of daily attendance provide a tool for economists and other education researchers to measure school "disengagement" and other theories of gradual high school dropout in a systematic way. Disengagement is the process through which students become less connected to school and potentially eventually leave (Christenson et. al., 2012; Rhumberger 2004 for overviews). Disengagement is difficult to quantify because it is an abstract concept. Empirical studies of the path of disengagement are dependent on the measures of disengagement they use. Objective measures such as assignment completion, truancy, and grades are employed, as are student survey responses to questions about how students feel about school. The disengagement literature, like the absence literature, focuses on yearly aggregates and correlations between attitudes measured at a single point in time and outcomes. However, disengagement is a dynamic process and dropout is only one of the potential consequences (Lee & Burkham, 1992; Rhumberger & Larson, 1998; Rhumberger 2003). Within year attendance patterns potentially provide a much deeper look into the process through which students disengage from school and the student behaviors that underly yearly attendance totals or student reported affinity for school. Economists have focused substantially less on disengagement than other education researchers, due to the measurement issues, and within-year attendance patterns provide a path through which economists can incorporate the lessons of disengagement theory into models of human capital development.

The goal of this paper is to describe the within year patterns of student attendance and establish basic facts about the way students attend high school in order to provide insight into the mechanisms through which attendance may affect achievement and how attendance fits into models of human capital investment, disengagement, and the education production function. I develop three measures of within-year absence intensity: the averaged monthly attendance rate, the probability that an absence is part of a multi-day spell and the length of the spell, and the persistence of absence which is the transition probability of absence to absence and absence to present. I examine how all three of these measures, as well as more traditional measurements of averaged attendance and chronic absenteeism, vary over the school year and type of absence. Truant absences are more likely to be part of multi-day spells of absence, they have a higher rate of persistence, and spells of truant absences are longer than spells of excused absences. I find that absences of all types are least common in the beginning of the year and that truant absences peak in June whereas excused absences peak between December and March. Over the course of the school year, truant absences become more common, more persistent, and are, on average, part of longer spells of truancy.

Like the previous literature, I find that aggregate attendance is highly correlated with academic success. However, given a total number of absences, the timing and type of absence also matter. Students with more truant absences have lower grades, lower test scores, are more likely to dropout than peers with the same number of total absences but fewer truancies. A student with a single truant absence in September of 9th grade is 10 percentage points less likely to continue into 10th grade in a traditional high school than a student who misses over a month of school with excused absences. High and low achieving schools and students also have very different patterns of absence and truancy accumulation over the year. Students in low performing schools are more likely to be truant than students at high performing schools and that gap increases over the course of the year. However, all schools have similar patterns of excused absences. While I can not distinguish between the causal effect of these attendance patterns on student achievement from unobserved student heterogeneity, the insight of these pattens provides valuable information to both researchers and policy makers on how students attend high school as well as early warning indicators for teachers and parents.

The paper proceeds as follows. First, I discuss the most common aggregate measures of student attendance used in the education and policy literature including averaged yearly attendance, chronic absenteeism, count days, and retrospective survey response. I follow that discussion with a description of the LUSD student-by-day level attendance data and discussion of how the characteristics of the data provide insight into high school attendance patterns that are impossible using aggregate attendance. In the methodology section, I discuss how different constructions and definitions of standard aggregate attendance measures affects the information they provide researchers. Following aggregate attendance, I develop three novel measures of attendance intensity based on the daily level data: monthly averaged attendance, the prevalence and length of multi-day absence spells, and the persistence of absences. Finally in section 4, I present estimates of these different measures of attendance, both aggregate and daily, discuss how the measures change over the school year, and their relationship to student achievement. I finish the section by discussing a number of early warning factors for academic achievement that show up during the first month of high school. I conclude the paper with a discussion of the relevance of these patterns to economic and education research as well as their policy implications.

1.2 Data

1.2.1 Standard Measures of School Attendance

There are three common aggregate measures of yearly attendance rates used in the education literature, accountability programs, and school funding formulas: averaged yearly attendance, chronic absenteeism or chronic truancy rates, and count days. It is important for both researchers and policy makers to understand what each of these statistics measures and how they are constructed. The daily attendance rate is the number of students attending school divided by the number of students registered at the school each day. Averaged yearly attendance is the average of all the daily attendance rates for the school year. Averaged yearly attendance is a required federal accountability measure.

Recently, there has been increased policy focus on attendance and chronic absenteeism as additional measurements of student outcomes and school quality. Unlike averaged yearly attendance, which measures how many of a school's students are in school on an average day, chronic absenteeism measures how many students are absent frequently during the school year. A student is considered a chronic absentee if she misses more than a specified number of days per year for either excused or truant absences. Researchers, policy makers, states, and the federal government have different cutoff levels for how many absences qualify as "chronic." When the Department of Education launched the Every Student, Every Day national initiative in conjunction with then President Obama's My Brother's Keeper (MBK) program, they defined chronic absenteeism as missing 15 or more days during the school year. Under the Every Student Succeeds Act (ESSA) chronic absenteeism, in addition to average attendance rates, has been added to the Civil Rights Data Collection requirements but there is no specific cutoff for what qualifies as chronic absenteeism named in the bill (US Department of Education, 2016). Many states including New Jersey, Hawaii, Oregon, California, Connecticut, and Georgia already use chronic absenteeism as a metric of school accountability or as part of their funding formula (Attendance Works, 2015). Most states set the chronic absentee cutoff at 10% of the school year (17 - 20 days depending on the length of the school year) and the 10% rule is the focus of most researchers and policy organizations (Attendance Works, 2015; Hough et al, 2016; Education Commission of the States, 2016; Allensworth & Easton, 2007).

An additional attendance measure is chronic truancy. Like chronic absenteeism, chronic truancy measures the number of students in a school who are truant for a substantial number of days rather than the average number of students who are truant on any one day. Truancy is any day in which a student is required to be in school, is not in school, and does not have a legally valid excuse for their absence. Students are considered chronic truants once they pass a certain level (cutoff) of truant absences in one year. LUSD's state defines chronic truancy as 3 or more truancies in a year.

A count day is a single day during the school year in which the students attending that day are counted by the state in order to assess enrollment and distribute funding. As of 2012, there were 19 states that used either one or multiple count days in order to assess enrollment and attendance and distribute funding to schools (Lara et al, 2012). How count days measure student attendance and enrollment depends on when the count day is during the year, and whether it only counts students who are present on that specific day or all students who are enrolled as of that day. Different states have count days at different times of the year but the vast majority choose a count day early in the school year and about half of states with count days have multiple count days.

Researchers also use retrospective survey data to measure attendance and the relationship between attendance and achievement. The National Education Longitudinal Study of 1988 (NELS: 88) and the Education Longitudinal Study of 2002 (ELS: 2002) both ask students retrospective questions about how often they missed school or skipped school during a past time period (4 weeks to 5 months) and parents are asked about whether they were contacted by their child's school about attendance.

1.2.2 LUSD Administrative Data

The measures of attendance discussed above provide aggregate information on different dimensions of school attendance. However, these measures of attendance provide little insight into how students attend school and which patterns of attendance are most predictive of academic achievement. Understanding students' patterns of attendance, specifically when they are missing school during the year, what they are missing school for, how absences are grouped over time, and how those patterns are correlated with student achievement is important for developing effective policies to address absenteeism and also to understand the role that attendance plays in the education production function.

I use extremely detailed day-by-student administrative attendance data to document within year patterns of student attendance and their relationship to measures of student achievement including grades, test scores, graduation, and dropout. Using this individual data, I am able to track how student's attendance behavior changes over the course of the school year, why students are missing school, how long students are out of school at a time, and how those patterns relate to student achievement.

LUSD provides an excellent case study of student attendance behavior. It is among the 50 largest districts in the country and has an averaged freshman graduation rate nearly at the mean of large districts (NCES, 2015).[1] LUSD is majority minority: 40% Hispanic, 10% Black, 27% White, and 20% Asian and, as Figure 1 demonstrates, there is large variation in school racial composition. This is because demographic composition of the school a student attends is determined by the neighborhood catchment area in which she resides. The within district racial and achievement diversity makes LUSD an ideal context to examine differences in attendance patterns among high and low performing schools and their relationship to student achievement. All students in the district are subject to the same graduation requirements and the same attendance rules and enforcement.

This paper links six administrative datasets within the school district: the student entrance file, student exit file, transcript information, accountability standardized test scores, high school exit exam dates and scores, and daily level administrative attendance records for five cohorts of 9th graders. I limit the sample to students enrolled in 9th grade in a "traditional" or neighborhood public high school all of which have the same attendance rules. Students are between the ages of 13.75 and 15.5,[2] are diploma bound, enter on-time during September of their freshman year, and do not leave before the end of the first

[1]Ranking and dropout rate are not reported in order to protect the district's anonymity
[2]This is consistent with state school entrance policy

month.[3] The sample contains over 25 million day-by-student observations for over 38,000 individual students.

The key dataset is the Final Attendance Code Database. This administrative data set includes one code per student per day while the student attends her original public high school. The daily attendance code is assigned by the district office. In each class, teachers mark the student absent or present and then the district assigns an attendance code based on the reason the student is out of class. Students who are absent from school fall into two major categories: excused absent and truant absent. Excused absences are those for which the student has a legally valid excuse such as an illness, religious holiday, or funeral. Truant absences are all other absences. A student is truant if she is not in school on a day in which she is legally required to be there and does not have a legally valid excuse for the absence. Students are required to be in school if they are under 18, do not have a GED or diploma, have not transferred to another school district, and are not home schooled. Students may still be considered truant even if their parents excuse them from class. Students are required to have documentation for illnesses lasting more than 3 days, and there are strict rules for the amount of time that can be taken for bereavement and religious holidays. Students who are in school can either be in school and on time (present) or in school and late (tardy).

Overall, most students are in school most days. On an average day 85% of students are in school and on time, 3.5% are out sick, 0.4% are out for a family engagement of some sort, 6% are tardy, and 2% are truant absent. Most absences from school are the result of reported illness or medical appointments. Over

[3]This is to eliminate students who were not actually attending the school but were registered in it as a holdover from the previous year. It is the same sample selection process as Wedenoja (2017)

half, 57% of absences are for medical reasons.

The attendance data are augmented by student demographic and achievement data. Demographic data comes from LUSD's student entrance file. The entrance file includes all students who enter into a traditional public high school between the years 2005 and 2009. Dropout and transfer outcomes are from the student exit file, which contains detailed administrative exit information generally unavailable in survey datasets. It distinguishes between types of transfers (in distinct, public in state, private in state, out of state, out of country) and types of diplomas that students earned. Diploma types include a high school diploma from a traditional school, a high school diploma from an alternative school, an adult education diploma, a GED, or no diploma. Alternative schools are high schools within LUSD that waive some of the requirements of traditional high schools. They can have non-standard school days, online learning components, lower credit requirements for graduation, and may waive some higher level classes, especially in math, that are otherwise required for graduation. Alternative schools do not include other schools of choice within the district such as magnet, gifted, or area specialty schools which have the same requirements as traditional neighborhood public schools. Diplomas from alternative schools are likely to have lower value than those from traditional public high schools due to their reduced requirements. Furthermore, nearly half of students in the sample who transfer to an alternative high school eventually drop out.

In addition to graduation and exit data, achievement data include students' high school exit exam scores, yearly weighted grade point averages,[4] and accountability test scores in math and language arts. Students in the same grade

[4]The weighted GPA uses a multiplier for advanced courses and is the one the district uses to determine if a student has passed the GPA threshold of 2.0 to graduate

all take the same language arts test but the math test is for the specific math class taken that year regardless of grade. The high school exit examination is required for graduation and first attempted and usually passed in the spring of tenth grade, 80% of 10th graders pass the exams before 11th grade but students who fail may continue to re-take the test until they pass.

While the data are extremely detailed, they are limited by parent honesty. Parents may lie about the reason behind their students' absences which could cause measurement error in the composition of absence types. While parents' truthfulness when providing the school an excuse does not alter if a student is in school or not, parental lying will cause error in analysis of absence type. However, parents do not have absolute power to excuse students from school. Parents can not give students unlimited excused absences and may only excuse students for illness, medical appointments, bereavement, documented family emergencies, and religious holidays. Beyond three consecutive days of "illness" or "medical appointments" parents must provide written documentation of the medical condition from a doctor. Students who are seriously ill also have the option to attend school remotely. Additionally, bereavement beyond a specified number of days (based on funeral location) or an undocumented family emergency beyond one day are labeled unexcused and truant despite a parent's excuse. While the reason a student is out of school is an important componenet of my anyalysis, I also examine patterns the time students spend out of school regardless of why they are absent. These patterns are not dependent on parents providing schools with the true reason for their student's absence.

1.3 Methodology

1.3.1 Measuring Chronic Absenteeism, Chronic Truancy, and Averaged Yearly Attendance

In addition to averaged yearly attendance, chronic absenteeism and chronic truancy are increasingly popular school accountability measures. However, there is no universal standard for how many absences or truancies a student must have during the year before she is considered to be a chronic absentee or truant. There are two important factors in choosing a cutoff for what qualifies as chronic absenteeism. First, there should be a meaningful difference in achievement between students below or above the chronic absentee cutoff and second, the measure should be able to distinguish schools that have problem attendance from those that do not. Most students miss some school during the year so if the cutoff for chronic absenteeism is too low, schools with severe absenteeism problems would be indistinguishable from schools with mild absenteeism If the cutoff is set too high, then it is not a meaningful indicator of student achievement. Using the detailed LUSD data, I examine how chronic absenteeism rates are affected by the number of absences chosen as the chronic absentee cutoff and how well those measures distinguish low and high performing schools from each other.[5] I focus on four chronic absentee cutoffs: 5%, 10%, 15%, and 20%. I do a similar exercise for chronic truancy, and examine how the chronic truancy rate changes when 1 to 5 truancies qualify a student as a chronic truant.

In addition to these measures of chronic absenteeism and truancy, I also ex-

[5]high and low performing schools are schools in the bottom and top quartiles by average test scores and graduation rate.

amine how different definitions of averaged yearly attendance and count days affect the calculation of student attendance and the differences in attendance between schools. To do so I define "present" in three ways: present at all on the day, present all day, and present and on time. I also show how the timing of a count day affects how absenteeism is measured by comparing a count day in September and a count day in June.

1.3.2 Measuring Within Year Patterns in Absence Composition, Timing, and Intensity

The main focus of this paper is to document the within year patterns of student attendance that underline the aggregate measures and provide better insight for economists, education researchers, empiricists, and policy makers. In this paper I focus on the patterns of attendance of 9th grade students. Attendance is viewed to be particularly important during the first year of high school as the transition from middle to high school is disruptive to students in terms of peer group, school change, and more difficult classes as well as increased autonomy over their choices. Studies of large urban school districts have found that students with lower attendance in 9th grade are less likely to graduate in four years and have lower grade point averages (Chicago: Allensworth & Easton, 2007; New Jersey: Advocates for Children of New Jersey, 2016; Baltimore: Olson, 2014; Utah: Utah Education Policy Center, 2012). Students are likely to establish habits of attendance during 9th grade that persist through their high school careers.

Beyond the aggregate measures of school attendance, I examine three main facets of absence patterns within the school year and their interaction: type of absences and absence excuses, the "intensity" of absences, and how type and intensity of absence vary across demographics, schools, and time. The main two absence types are excused absences, including sick days and certain family obligations, and truant absences.

I construct three intensity measures for within year absences. The first is the averaged monthly attendance rate. This is similar to the averaged yearly attendance measure required for accountability and is simply the monthly average of the daily attendance rate. The goal of this measure is to document how average attendance changes over the course of the school year. Attendance may have seasonal patterns due to holidays, weather, exam schedules, and teen work opportunities. The second is the probability that an absence is part of a multi-day spell of absences. An absence on day t is part of a multi-day spell if the student is also absent on day $t-1$ or $t+1$. If a student is absent on a Friday and also the following Monday but returns to school on Tuesday, the Friday and Monday are part of a 2 day absence spell. Additionally, I document the average length of absence spells. Finally, I construct a measure of the persistence of absence. This is the estimated transition probability of absence to absence and absence to present by absence type. This allows me to look at both how absences persist but also how absence types interact with each other by measuring the transition probability from one type of absence, such as out sick, to another, such as truant.

Using these measures of absence levels and intensity, I document how they change over the course of the school year and the differences in these measures between schools and demographic groups. I break the district schools into four quartiles based on their graduation rate, average GPA, and average test scores.[6]

[6] The achievement measure chose to define the quartiles does not affect which schools are in

1.3.3 Measuring the Relationship between Attendance Patterns and Student Achievement

After documenting the basic patterns of student attendance measures including aggregate attendance, and attendance intensity, I relate those patterns to measures of high school achievement. High performing students and schools, on average, have substantially different attendance patterns compared to low performing students. The measures of achievement I focus on are eventual graduation, continuing from 9th to 10th grade, passing the high school exit exam, standardized test scores in English and math, and grade point average. High school attendance is also a non-cognitive skill outcome in its own right. The ability to regularly attend school on time with limited supervision is a valuable life and workplace skill that students develop in high school.

Documenting the observed relationship of attendance to measures of cognitive achievement provides valuable insight into the role of attendance in the education production function as well as insight into the mechanisms through which policy can affect attendance and attendance can affect achievement. The positive relationship between average attendance and student outcomes is well documented, however, evidence on the mechanisms through which attendance and outcomes are related is sparse. I establish novel correlations of attendance patterns and student achievement. I provide the first documentation of the relationship between absence intensity and achievement as well as the first look at how absence accumulation patterns and timing over the school year differ by outcome measures. High achieving schools and students have very different attendance patterns from their lower achieving counterparts. While I can

---

which quartile.

not distinguish whether this is a causal effect of the attendance patterns or driven by unobserved student heterogeneity, documenting these patterns will help schools direct academic interventions earlier.

1.4 How Students Attend High School

1.4.1 Average Attendance Rates

The rate of high school attendance when averaged across all students, all schools, and all days is very high. On average, 85% of students in the sample are present and on time, 6% are tardy, 4.2% have excused absences and 2% are truant. However, those low average rates of absence obscure substantial variation in attendance across students within schools and across schools. Figure 2 contains histograms of total days missed for all 9th graders within each school. Each panel is a different school. Some schools (2 and 13) have very few students with more than two or three absences, while other schools (8, 13, and 16) have far more students in the right tail who miss large amounts of the school year. The differences across schools in attendance are even more stark in Figure 3. Schools in the four quartiles of the academic achievement distribution do not differ greatly in the distribution of excused absences (bottom panel). However, students in the lowest performing schools have far more truant absences on average, and students in the 75th and 99th percentiles of the schools have over double the number of truancies as students in the 75th and 99th percentiles of high performing schools.

Table 1 presents versions of the standard aggregate measures of student attendance discussed above: chronic absenteeism, chronic truancy, averaged yearly attendance, and count days. The table contains aggregate attendance for all schools, and for schools in each achievement quartile. As the bottom panel demonstrates, how the averaged yearly attendance rate is calculated matters for comparing schools to each other and distinguishing which schools have attendance problems. If attendance is measured as the percent of students who are in school for any part of the day, low and high performing schools look very similar with attendance rates ranging from 94% to 97%. The lowest performing schools fall only one percentage point below the district average. However, alternative definitions of averaged yearly attendance reveal substantial differences in attendance at low and high performing schools. In the lowest performing schools (bottom quarter) only 76% of students are in school and on time compared to 90% of students at higher achieving schools. While students at low achieving schools are almost as likely to be in school for some of the day as students in higher achieving schools, they are much less likely to be on time or in school for the entire day. All quartiles of schools have similar count day attendance if the count day is at the beginning of the year (97% - 98%). However, lower performing schools have lower counts than high performing schools if the count day is at the end of the year falling from 97% attendance to 92%, whereas high performing schools only fall from 98% to 97%.

Schools that have similar averaged yearly attendance rates can still have very different chronic absentee and truancy rates. The goal of incorporating chronic absenteeism and truancy into accountability measures is to identify schools that have students missing large percentages of school. The rate of chronic absenteeism is the percentage of students who miss more than a certain cutoff number (or percent) of school days. Rates of chronic absenteeism and truancy are sensitive to how they are defined as shown in Table 1. If the cutoff for chronic absenteeism is set very low, 5% of the school year, students in the bottom, low, and middle quartiles have very similar chronic absentee rates ranging form 48% to 43%. The higher the threshold for chronic absenteeism, the more different low and high performing schools appear. Using the common chronic absentee cutoff of missing 10% of the school year, low performing schools have a chronic truancy rate of 24% and high performing schools have a rate of only 11%. Because low performing schools tend to have long right tails in the absentee distribution, the higher the cutoff for chronic absenteeism the larger the difference between high and low performing schools appears to be.

The difference in the chronic truancy rate across school quartiles is even more stark. Low performing schools have much higher rates of chronic truancy compared to higher performing schools even if chronic truancy is defined as a single truant absence. Nearly 70% of students in the bottom quartile of schools will be truant at least one day compared to only 30% of students in the highest quartile of schools. 28% of students in the lowest performing schools meet the district definition of a chronic truant by having 3 or more truant absences compared to only 11% of students in the highest performing schools. Table 2 breaks down rates and levels of excused absence, truancy, chronic absence, and chronic truancy as well as average days missed for individual schools. In the school with the lowest absence rate students only miss 6 days of school per year on average compared to 16 days on average at the school with the lowest attendance rate.

Absenteeism and truancy rates also vary across demographic groups as seen in Table 3. Black, Hispanic, and English Language Learner (ELL) students have the highest rates of truancy and absence missing 12 to 14 days on average compared to 10 days for white students and 5 days for Asian students. Nearly one third of Black students and over one third of Hispanic students are chronic truants and the chronic truancy rate is slightly higher for men, although they miss the same amount of days on average as women.

1.4.2 Attendance Intensity

Table 4 contains the measures of absence intensity. Absence intensity is measured as the percent of absences in multi-day absence spells and the length of those spells. On average, 42% of absences are part of a multi-day spell and the average spell length, given that a spell is at least 2 days, is 3.8 days with 13% of absences part of a spell that lasts a week or longer. Truant spells are longer than excused absence spells at 5.7. Truant spells are also much more likely to be a week or longer, 23% compared to 9.4% and just over half of truancies are part of a multi-day spell.

Another way to view attendance intensity is through the day to day transition probability between present and different types of absences. The four panels of Table 5 contain transition probability matrices for students in quartiles of student performance. Across all schools and absence types it is very unusual for students to transition from one type of absence to another, it does not appear to be the case that students are out of school for the three sick days they are allowed without documentation, and then continue to be truant absent. Only 5% of students out sick in the lowest performing schools transition to truancy, and only around 1% of students in higher performing schools. Transition probabilities from and to out sick and from out on disciplinary grounds are very similar across school performance quantiles. This is likely due in part to the strict rules that govern how long a student can be out sick without documentation and standardized suspension policies. However, in lower performing schools students who are truant are much more likely to stay truant. The truant to truant probability is 38% for low performing schools and 27% for the highest performing schools. Students in high performing schools are 17% more likely to return to school after a truancy than students in low performing schools.

1.4.3 Yearly Patterns

Attendance levels, type, and intensity vary substantially throughout the year. Table 6 shows that average attendance decreases over the year with the highest levels in September (96.6%) and the lowest levels in June (93.5%). Most of this decrease is due to an increase in truant absences from less than 1% in September to 3.2% in June. Monthly averages of truancy increase over the school year, whereas excused absences increase and peak in December and then gradually decrease. The peak in excused absences in December is likely due to parents pulling their children out of school early for winter break. In fact, the top panel of Figure 4 shows spikes in all types of absences around winter break and then a decline when students return in January.

Attendance intensity also varies over the school year as can be seen in Table 6. The percentage of absences and truancies that are part of a multi-day spell are lowest in September at 37% and 41% respectively, as are the own transition probability for truant (28%) and excused absences (25%). Total absences and absence intensity increase consistently over the school year with the absence rate nearly doubling by June and the truancy rate more than tripling. The percentage of absences and truancies in multi-day spells also increases to 50% of total absences and 41% of truancies. Students also become more likely to stay truant with 54% of students staying truant up from 41%. The change in intensity of total absences appears to be almost completely driven by truancy and both the level and intensity of excused absences peak in the middle of the year not at the end. Students are most likely to be out sick between December and March, which would be consistent with flu season and the excused to excused transition probabilities are highest during those same months. Sick days also have a weaker weekly pattern than truancies although students are most likely to be out sick on Mondays.

The difference in transition probabilities between high and low performing schools discussed earlier, results in high and low performing schools having yearly attendance patterns that look very different as shown in Figure 5. Low performing schools have a higher average level of absences as well as more intense spikes as seen in the top panel of figure 5. Most of that variation is driven by truancy (bottom panel) rather than excused absences (middle panel), the patterns of which are very similar for all schools. Low performing schools not only have higher truancy every day than higher performing schools, but they have much more intense spikes in truancy around school breaks and the truancy rate increases dramatically over the year relative to all other schools. One other difference between high and low performing schools is in disciplinary absences and family emergency absences. Figure 6 shows that students in low performing schools are more likely to be out on disciplinary grounds, but the pattern doesn't differ dramatically from other schools. For family emergencies, low performing schools have much higher spikes near winter break and the end of the year.

1.4.4 Attendance and Student Achievement

Students who are absent more often have worse academic achievement than students who are absent less. While this is a well established fact in the literature, the relationship between the more detailed measures of attendance rates and intensity to academic achievement has not been documented before. Figure 7 plots the relationship between total days missed freshman year and eventually graduating with a traditional diploma (top panel) and continuing in traditional school in 10th grade (bottom panel). Each circle in the scatter plot is a school by number of days missed observation and is scaled in size by the number of students in that cell. Vertical reference lines are two different measures of chronic absenteeism: a 5% and 10% of days missed cutoff.

There is a clear negative relationship between days missed freshman year, continuing on into 10th grade, and eventually graduating. However, while students who miss more than 10% of days are less likely to continue school, there is also a difference between students who miss more than 5% of days and those who miss fewer. Table 7 contains average continuation and dropout rates by total number of missed days. Students who miss 18-22 days of school freshman year (about 10%) are 10% less likely to continue in traditional school than students who miss 13-17 days and 17% less likely to continue than students who miss fewer than 3 days. They are also much more likely to dropout. The relationship between between school days missed and continuing school is weaker for students who have low or no truancies; 81% of students who have fewer than 3 truancies and miss 18-22 days of school continue in traditional school which is just under the average of 82%. If students have no truancies, 86% of those who miss 10% of the school year continue which is above average. The apparent relationship between dropout and chronic absence appears to be driven mostly by truants.

As Table 8 shows, measures of chronic absence and truancy do predict whether students will continue on in traditional school for all grades. In 9th and 10th grade chronic absentee students are 11 and 9 percentage points less likely to continue school than students who are not chronically absent, provided they are not also chronic truants. Chronic truants who miss more than 10% of the school year are 30 percentage points less likely to continue to 10th grade and 36 percentage points less likely to continue to 11th. Chronic absentee students are only 2 percentage points less likely to continue to 12 grade after making it to 11th. However for students who are also chronic truants, they are still more than 30 percentage points less likely to continue.[7]

Students who continue in traditional school also have different within-year patterns of attendance seen in Figure 8. Students who transfer into alternative schools are more likely to be out sick early in the year than students who stay in traditional school, they are also much more likely to have disciplinary absences. The middle panel of Figure 8 shows that students who drop out between 9th and 10th grade have an especially stark pattern of increasing truancy over the course of the year compared to students who continue in any kind of school.

Measures of academic achievement and cognitive skills also differ by at-

[7]Students dropout and transfer between years so the students who are still in traditional school in later grades are higher achieving than other students.

tendance level and intensity. Figure 9 shows that students with fewer missed days freshman year have higher GPAs and higher ELA test scores and Figure 10 shows that the pattern holds for passing the ELA and Math exit exams as well. The daily patterns of attendance also differ by academic achievement. Students with low GPAs have higher rates of absence, their absence rate increases more rapidly over the year, and that increase is driven almost entirely due to increased truancy as seen in Figure 11. Figure 12 shows that exit exam success follows a similar pattern, students who never pass their exit exams have higher rates of absence which increase over the year and that increase is also due almost entirely to truancy. Students who pass both exams the first time have a similar pattern of excused absences but virtually no truancy.

1.4.5 Early Warning Factors

Understanding the within year patterns of high school attendance can help identify early warning indicators for student achievement so that schools can intervene with students early during 9th grade and prevent patterns of bad attendance from becoming entrenched. September is the month with the highest rates of attendance, lowest truancies, fewest absences in multi-day spells, shortest average multi-day spells, and lowest absence to absence transition probabilities. However, despite those positive attributes, attendance behavior in September is correlated with attendance behavior later in the year and subsequent negative outcomes. Even small attendance issues in 9th grade could become larger problems later in high school. As such, attendance behavior in September, specifically the number and type of absences, can provide a valuable early warning sign for teachers and school administrators.

Table 9 describes the relationship between absences during September of 9th grade and outcomes later in the year and later in high school. As shown in panel A, missing even one day of September decreases the probability of graduating, GPA, standardized test scores, the probability of passing exit exams the first time, and increases the probability of dropout, having a GPA below the graduation cutoff of 2.0, never passing exit exams, and repeating Math. However, whether or not that absence is excused or truant matters. The difference between full attendance and one absence virtually disappears if that absence is an excused absence, and there are no truancies that month. Panel D shows virtually no difference in graduation or achievement for students who have one absence in September compared to no absences as long as that absence is excused.

A student taking a single sick day in September is not a good indicator of negative outcomes, however a single truancy is. Panel B contains the average achievement of students with 0-5 truancies during September of freshman year. Students with one truancy in September compared to none are 12 percentage points less likely to stay in traditional school the following year, more than twice as likely to eventually dropout, have an average GPA a full grade point lower, are substantially less likely to pass their exit exams, and have standardized test scores in ELA and math 0.6 and .17 standard deviations lower than students with no truancies. Nearly half of students who become chronic truants (miss 3 or more days) during their first month of high school do not continue on to 10th grade in traditional school and less than 20% will graduate with a traditional high school diploma.

Any level of truancy in the first month of high school is highly correlated with negative outcomes later in the student's career. A student who misses five days during September for excused absences has, on average, higher test scores in both English and Math than a student who is truant for a single day. The importance of even small numbers of truancies and absences in September highlights the value of student-by-day level data in designing effective dropout prevention policies. Schools can use these early warning indicators to target students early in high school in an attempt to prevent patterns of bad attendance and truancy from becoming permanent. Younger students are also more likely to be receptive to parent involvement. The importance of truancy in September suggests that districts should involve parents immediately upon a student's first truancy rather than waiting for the student to be labeled a chronic truant.

1.5 Implications For Economic and Education Research

1.5.1 The Becker Model of Human Capital Accumulation

The discussion above shows that much of the observed relationship between absence and achievement appears driven by truancy and that truancies are highly predictive of negative educational outcomes even at low levels of total absence. Recent expansions of the Becker human capital investment model have focused on sequential education investment in which students make the decision, each year, to begin another year of schooling based on their perceptions of whether or not they will be able to finish the year (Cameron & Heckman, 1998,2001; Cunha&Heckman 2007, 2008, 2010; Eckstein&Wolpin, 1999). The relationship between daily level, within year attendance dynamics and achievement allows those models to be further expanded to include within year sequential investment decisions in addition to the between year decisions.[8] It also points to the importance of student, rather than parent, decision making as early as 9th grade when students are only 14 to 15 years old. This is because truancy, rather than excused absence, has a stronger link to student outcomes, and truancy is a student decision.

In its traditional form, the Becker model relies on actors to be rational, forward looking agents who are capable of making predictions about the effect of their current education investment decisions on future outcomes. They are also able to follow through on those decisions. This may not be a reasonable assumption for young teenagers who decide whether or not to be truant. Teenagers are less able than adults to understand the future consequences of their actions and they are hyper-sensitive to immediate rewards (Blakewell & Robbins, 2008). Progressively worsening spells of attendance are not consistent with a standard Becker model which predicts that students should attend school all at once because they will attend until the marginal benefits equal the marginal costs. When a student misses a day or week of school, it increases the marginal cost for future days because the student will have to catch up to the class or her skills may depreciate while out of the classroom.

In addition to providing a framework of expanding the Becker model to incorporate within year decisions, the patterns above indicate that relaxing assumptions of student rationality in the investment model may improve its ability to explain the persistently high dropout rates in the US. There have been a number of studies of education that incorporate principles from behavioral economics and psychology. [9] Students in this sample appear to be highly sensitive

[8] A version of a daily investment model is discussed in Wedenoja (2017).

[9] See Lavecchia et al (2014) for an overview.

to small changes in the psychic cost of a day of school which is not consistent with an agent maximizing lifetime expected utility. Students are most likely to be truant on Monday when they are likely tired from the weekend and out of the habit of getting up on time and are more likely to miss days around school breaks potentially because they are anticipating not being in school. Levitt et al. (2016) finds similar short term incentive effects on student effort. When students face a financial incentive for scores on a low stakes test, they put in effort and achieve higher scores than when they are not offered an immediate incentive.

One of the patterns observed in this data is that some students miss large amounts of school, a month or more, and then still return to school the following year to try again. This could be evidence of naive present biased preferences (O'Donoghue & Rabin; 1999). Gradually worsening attendance patterns followed by a return to school the following year, as evidenced in the data, is not consistent with exponential discounting as used in the Becker model but is consistent with a naive present biased agent who begins the school year with a plan of attendance and believes that her future self will follow through with that plan. However, once the day the student plans to attend is the present, the student will decide that the cost of school is too high and not attend despite her original plans. She may then repeat this process the following year because she does not recognize her own time inconsistency.

1.5.2 The Education Production Function

It is ex-ante unclear how student attendance should enter into the education production function. It can be used as an output to a production function in which the inputs are school, student, and parent resources. As an output it has been used as a measure of disengagement from school (Astone, 1991; Archambault et al, 2009) and a measure of non-cognitive skills (Imberman, 2011). It could also be an input as an investment of student time and effort into a production function that has education achievement as the output. As an input, attendance has been used in the education production function as proxy for student effort and engagement (Lamdin, 1996) and in conjunction with instructional time (Coates, 2003; Macerotte, 2007).

The description of attendance patterns above supports the idea that attendance is an important input into the education production function. The type of absences that are most correlated with negative outcomes are truant absences, which are choices that students make over their education investment. That is not to say that school policy and inputs do not have an affect on student attendance. The patterns above also provide insight into how the timing and type of absences enter into the education production function given a total number of absences.

Students do not attend school all at once. Over time, students with worse outcomes have more frequent spells of truancy and those spells increase in length. Because of this, aggregate attendance does not appear to be the complete picture of how attendance enters into the education production function. Only 84% of students who miss more than 10% of the year will continue to 10th grade compared to 88% of students who miss less than that, and that number is lower (63%) for students who have had some of their absences as a multi-day spell, and even lower (54%) if any of those absences were part of a multi-day truancy streak. Even students who only miss 5% of the school year are less likely to continue if any of those absences are part of a multi-day spell. This is suggestive evidence that student skills may depreciate during the school year when, given the same total number of absences, students are out of school for longer consecutive periods of time. There is substantial evidence that student skills depreciate over the summer (Cooper et. al, 1996; Downey et. al, 2004; Alexander et. al, 2007) and worker's skills, including critical reading, depreciate while they are out of work (Arulapalam et al, 2001; Edin & Gustavsson, 2008). The results above suggest that skill depreciation may also happen during the school year. The grouping of absences also matters for test scores and achievement. Given a total number of absences, having some of them as part of a spell reduces test scores in math by an additional 0.1-0.2 standard deviations and 0.12 - 0.24 standard deviations in English. If those are spells of truant absence, ELA test scores fall by a further 0.3 standard deviations and math by a further 0.2 standard deviations. Additionally, test scores and continuance are lowest for students with longer individual spells out of school and for students who have ever missed more than a week of school at a time. These patterns are consistent with incorporating attendance spells, in addition to total attendance, into the education production function. Using total attendance implicitly assumes that all absences are the same in how they contribute to student achievement during the year. Multi-day spells have a larger affect on achievement, which could be explained by skills, especially newly learned skills, depreciating and being forgotten faster during longer absences.

Finally, within year attendance presents another mechanism through which students' peers affect their educational attainment. When students are absent frequently and for long periods of time they fall behind their more regularly attending peers. One likely consequence is that teachers have to spend more time with these students to try and catch them up to the rest of the class, or teach at a slower rate so that they do not fall so far behind that they can not catch up. Students with a higher fraction of absent peers may receive less direct instructional time than students with a lower fraction of absent peers. An alternative possibility is that teachers do not take into account chronic absentee students and do not spend extra time with them. Although that may be better for the achievement of present students, it could actually accelerate the process of school disengagement as absentee students fall further and further behind their peers potentially dropping out when it becomes impossible to catch up. Absent peers could also affect other students' decisions to attend school. Students may feel pressured to spend time with peers out of school, or enjoy school less when their friends are not present.

1.5.3 Empirical Implications of Attendance Measures and Patterns

Attendance is commonly used as a desirable outcome in policy analytic studies as well as in state and federal accountability scores. The discussion above stresses the importance of researchers and policy makers who use measures of attendance and absence to know exactly how they are constructed and what they are measuring. Both high and low performing high schools have similar averaged attendance levels when students are counted as present if they are in school at all during the day but, when present is defined as in school and on time, low performing schools have much lower rates of averaged yearly attendance. Policies targeted at improving attendance that function mainly through encouraging students to be at school first thing in the morning, such as free breakfast, or encourage students to be in school for an entire day would appear to have little effect if only an averaged yearly attendance rate is used. In fact, studies of free breakfast find no significant effect on averaged yearly attendance but most can not measure whether the policy increased the amount of the school day that students spent in school (Leos-Urbel et al, 2013). Similarly after-school programs that encourage students to stay through the end of they day would also have muted measured effects if they encourage attending students to spend more time in school rather than enticing out of school students back.

Focusing on average attendance rates also misses students who have serious attendance problems. Chronic absenteeism is a better measure of the severity of attendance problems within a school. However, how chronic absenteeism is defined matters. When the cutoff for chronic absenteeism is low, missing 5% of days, high and low performing schools appear more similar, which hides the fact that lower performing schools have substantially more students who are missing 10% or even 20% of school days. Single cutoffs for chronic absenteeism could also create unintended negative incentives for school districts. If the chronic absenteeism rate is the percentage of students who miss more than 10% of the year, about 18 days, then the school has an incentive to keep students from missing that 18th day but very little incentive to prevent students from missing any subsequent days. This is especially troubling because the data show that there is no discrete jump in achievement after missing 5% or 10% of the school year. Every day that a student misses matters for their ability to graduate and their achievement on tests. Researchers using measures of chronic absenteeism should make sure they know exactly what cutoff the school or state uses. Programs that target students who have the worst attendance records, such as high intensity mentorship or social worker involvement, may reduce absences among their target students but even if those students increase attendance if they are still above the cutoff for chronic absenteeism, then the policy will appear to be no effect on chronic absenteeism.

1.6 Conclusion and Policy Implications

Recent changes in education reporting requirements at the federal and state level now include rates of chronic absenteeism in addition to average measures of attendance. Many school districts and non-profit organizations, such as Attendance Works and City Year, have recognized the importance of attendance and chronic absenteeism for student success and have supported programs to try and get students to school. Attempts to decrease both daily absentee rates and chronic absenteeism include policies such as school and family partnerships and school reforms to intervene with students who have attendance problems (Epstein & Sheldon, 2002; Balfanz *et. al.*, 2007; Balfanz & Byrnes, 2014). The main attributes these programs have in common is identifying students early and connecting them with high intensity mentors both to help them with their schoolwork and push them to attend class.

The discussion above reveal some limitations to the current policy discourse on the importance of attendance. The recent focus on chronic absenteeism has been an important step forward in understanding how attendance interacts with student achievement. However, the focus on keeping students' absence rate below 10% may be too high of an acceptable threshold and, thus, not the best goal. Student achievement decreases for any absences beyond one or two and students start to fall below average in graduation, test scores, and GPA after missing 5% of school days, not 10%. Focusing on 10% as the only important threshold does not fully address the importance of other absences.

Additionally, the data show that it is truant absences, not all absences, that are most predictive of achievement. Students who miss 10% of the school year for excused absences still have test scores, school continuation, and GPAs around the average for the district but the more truant absences a student has, the lower her achievement is. Nearly half of students who miss more than 10% of days with 3 or more truancies will not continue in traditional school from 9th to 10th grade, compared to only 23% of students who miss 10% of days and have fewer than 3 truancies.

Attendance risk factors show up as early as September of 9th grade. Having a single truancy in September of 9th grade is more predictive of not staying in traditional school than missing 20% of the school year for only excused absences. Policy makers should focus on policies designed specifically to prevent truant absences, and to intervene with truants earlier to address the reasons behind truancy. The most effective truancy prevention programs address both attendance and academic problems. High intensity mentoring has reliably reduced truancy and increased test scores for at-risk students even late in high school (Bell et. al., 2015; Oreopoulos et al, 2014; DeSocio et al, 2007; Marburger, 2006; Lemieux et. al, 2011). The importance of truancy also suggests that schools should intervene with truants immediately after the first truancy, rather than waiting until the third truancy to refer students to attendance remediation as is the policy in LUSD.

Taken together, the results above provide valuable insight into the relationship between daily student attendance and education outcomes. Students who have more absences in 9th grade are less likely to graduate, less likely to remain in traditional public school, more likely to dropout, have lower grades, lower test scores, and are less likely to pass their high school exit examinations. These outcomes are the lowest for students with more truant absences, and students who have more frequent and longer multi-day absence spells. Attendance decreases over the school year and shows significant seasonal and day of week pattens.

1.7 Tables

Table 1.1: Absence Measures by School Performance

Absence Measure	All Scools	Bottom Quartile	Low Quartile	Middle Quartile	High Quartile
Absent>5%	42%	48%	43%	46%	36%
Absent>10%	17%	24%	18%	17%	11%
Absent>15%	8.4%	13%	9%	7%	3%
Absent>20%	4.6%	8%	5%	3%	1%
Truancies>0	49%	69%	47%	39%	30%
Truancies>1	35%	55%	31%	24%	17%
Truancies>2	28%	45%	24%	18%	11%
Truancies>3	22%	39%	19%	14%	8%
Truancies>4	19%	34%	15%	11%	6%
Truant>5%	11%	21%	8%	6%	2%
Truant>10%	5.4%	11%	4%	2%	.8%
Truant>15%	3.2%	7%	2%	1%	.3%
Truant>20%	2.0%	4%	1%	.8%	.2%
% Preset at all	95%	94%	95%	95%	97%
% Present all day	91%	86%	90%	91%	94%
% Present and on time	84%	76%	85%	85%	90%
% Count day in September	97%	97%	97%	97%	98%
% Count day in June	95%	92%	95%	96%	97%

Absent>n% is the percent of students who miss more than n% of the school year for any reason. Truant>n is the percent of students who have more than n truancies, and Truant>n% is the percent of students who are truant for n% of the year or more. Present at all is the average attendance for the school year if students are counted as present if they attend any part of the school day. Present all day is the average attendance for the school year if students are only counted as present if they attend the entire school day, Present and on time is the average attendance for the school year if students are only counted as present if they attend the entire school day and arrive to school on time. All schools is the average measure for all LUSD. Bottom, low, middle, and high are then average measures for students in the bottom to top quartiles of school performance on accountability tests. All measurements are for Freshmen.

Table 1.2: Attendance by School

Initial School	Total Students	all absences	excused absences	truant absences	chronic absentee	chronic truant	never pass exit exam	mean GPA
1	1,800	11.27	7.74	3.53	17.5%	24.5%	3.8 %	2.2
2	1,650	14.35	7.06	7.28	25.4%	39.7%	10.9%	2.1
3	3,200	8.400	6.98	1.41	9.56%	9.80%	2.2 %	2.5
4	2,570	15.91	5.21	10.7	28.0%	50.8%	9.9%	2.0
5	2,300	8.940	6.62	2.32	11.0%	15.6%	3.6%	2.3
6	2,000	8.150	7.10	1.05	8.00%	8.60%	1.2%	2.9
7	1,470	13.89	7.10	6.78	22.3%	40.8%	9.6%	2.1
8	1,540	15.97	11.4	4.58	29.6%	29.2%	5.7%	2.4
9	3,200	5.970	4.44	1.52	6.00%	10.5%	1.6%	2.7
10	2,100	12.04	7.78	4.25	18.5%	28.5%	6.5%	2.3
11	2,500	10.50	5.51	4.98	14.5%	27.6%	5.6%	2.3
12	2,600	11.10	8.87	2.22	16.7%	15.6%	3.0%	2.5
13	3,170	15.11	6.76	8.35	25.8%	44.3%	9.7%	2.2
14	3,000	8.310	6.93	1.37	9.50%	9.60%	1.0%	2.9
15	2,730	8.170	5.94	2.24	9.90%	15.2%	3.6%	2.6
16	2,300	10.36	9.14	1.21	15.3 %	7.50%	1.1%	2.9
total	38,000	10.81	6.97	3.83	15.88%	22.57%	4.6 %	2.5

Total students is the rounded number of students enrolled in the school for 9th grade in the sample. All absences, excused absences, and truant absences are the average number of all, excused, and truant absences respectively for each school Chronic absentee is the percentage of students in the school who miss more than 10% of the school year for any reason, chronic truant is the percentage of students who are truant for 3 or more days, never pass exit exam is the percentage of students in the school who do not pass their high school exit exams and mean GPA is the average GPA for freshman at each school.

Table 1.3: Attendance by Demographics

Demographic Variables	Sample Mean	Chronic Absence	Chronic Truant	Total Days Missed	Total Days Truant
Black	.117	20.6%	31.8%	12.7	5.2
White	.252	13.3%	8.50%	10.1	1.5
Hispanic	.440	21.3%	33.6%	12.7	5.7
Asian	.172	5.8%	7.80%	5.30	1.5
Female	.485	16.6%	22.0%	11.0	3.7
Male	.515	15.2%	23.1%	10.5	4.0
ELL		25.3%	43.0%	14.0	7.8
Black Female		20.7%	30.8%	13.1	5.2
Black Male		18.6%	32.8%	12.3	5.2
White Female		13.3%	8.50%	10.4	1.4
White Male		12.2%	9.90%	9.80	1.6
Hisp. Female		21.4%	32.9%	12.9	5.5
Hisp. Male		19.8%	34.1%	12.5	6.0
Asian Female		5.80%	7.80%	5.50	1.3
Asian Male		5.30%	8.10%	5.20	1.6

Averages are calculated off the 9th grade initial sample. Racial categories are coded by the district, Hispanic, Black, White, and Asian are mutually exclusive categories. Traditional graduates are those that receive a diploma from their original school, traditional dropouts drop out without transferring first, alternative graduates transfer then graduate, alternative dropouts transfer then drop out, transfer out means the student leaves the district for another.

Table 1.4: Absence Intensity Measures by Type of Absence

	% of school days	% of absent days	% absences in multi-day spell	Average Spell Length	% Absences in 5 day+ spell
All Absences	6.0%	—	42.6%	3.8 days	13.1%
All Excused Absences	4.0%	64.5%	38.9%	2.8 days	9.40%
Truant	2.1%	35.0%	50.1%	5.7 days	23.0%
Out Sick	3.5%	57.0%			
Family Engagement	0.4%	6.30%			

The table contains five measures of absence intensity for all absences, unexcused absences, and truant absences and two measures for the subcategories of excused: sick and family engagement. A multi-day spell is any group of two or more sequential absences of the same type so a truant streak only involves sequential truant absences but an absence streak is any type of absence. The average streak length is conditional on a multi-day spell and absences in a 5 day+ spell is the percentage of the absence type which occur during a group of a week or more consecutive absence. All observations are freshman students.

Table 1.5: Absence Transition Probabilities by School Performance

Panel A: Bottom Schools

t \t+1	present	sick	family	bus	disciplinary	truant
present	94.24	2.50	0.24	0.01	0.13	2.88
sick	73.08	21.74	0.55	0.03	0.23	4.38
family	49.01	3.37	43.29	0.03	0.19	4.12
bus	80.00	5.00	0.00	0.00	0.00	0.00
disciplinary	33.55	1.57	1.06	0.06	57.36	6.40
truant	59.41	2.17	0.27	0.03	0.29	37.83

Panel B: Low Schools

t \t+1	present	sick	family	bus	disciplinary	truant
present	95.37	3.06	0.01	0.01	0.10	1.25
sick	74.88	22.86	0.00	0.01	0.17	1.71
family	52.39	4.39	0.02	0.14	0.22	2.28
bus	72.31	2.31	22.31	0.00	0.00	2.31
disciplinary	37.71	1.82	0.00	0.11	56.35	2.82
truant	63.15	2.25	0.01	0.01	0.28	34.04

Panel C: Middle Schools

t \t+1	present	sick	family	bus	disciplinary	truant
present	95.59	3.22	0.29	0.01	0.09	0.79
sick	77.28	21.22	0.36	0.02	0.06	1.05
family	57.49	3.19	37.16	0.09	0.14	1.92
bus	92.36	2.87	0.64	0.00	0.32	2.55
disciplinary	33.72	1.47	0.95	0.03	61.89	1.93
truant	66.63	2.46	0.35	0.02	0.28	30.23

Panel D: Top Schools

t \t+1	present	sick	family	bus	disciplinary	truant
present	96.70	2.58	0.20	0.01	0.06	0.45
sick	77.53	21.32	0.26	0.01	0.06	0.80
family	59.02	3.69	36.42	0.08	0.02	0.77
bus	91.01	3.37	0.00	1.12	0.00	2.25
disciplinary	37.56	1.60	0.77	0.00	57.95	2.13
truant	70.57	2.06	0.18	0.03	0.26	26.88

Cells are the estimated day to day transition probabilities between attendance and absence types calculated for freshman. Panel A contains the probabilities estimated for students in the worst performing high schools, and panel D for the best performing high schools. The table reads left to right so the transition probability from present on date t to sick on date $t+1$ is the cell in the first row, second column of the transition matrix. Own transition probabilities are on the matrix diagonal.

Table 1.6: Attendance Characteristics of School Month and Weekday

	Absent	Excused	Truant	Out Sick	% Absence in Spell	% Truant in Spell	Truant to Truant Transition Prob.	Excused to Excused Transition Prob.
Panel A: School Month								
September	3.4%	2.6%	0.9%	2.3%	37%	41%	28%	25%
October	5.0%	3.6%	1.4%	3.3%	41%	43%	30%	21%
November	5.7%	3.9%	1.7%	3.6%	45%	49%	35%	23%
December	6.9%	4.6%	2.2%	4.1%	48%	50%	38%	25%
January	6.7%	4.3%	2.3%	3.8%	48%	52%	38%	28%
February	6.2%	4.1%	2.0%	3.8%	46%	50%	37%	27%
March	7.0%	4.4%	2.5%	4.0%	47%	51%	39%	26%
April	6.7%	4.1%	2.6%	3.6%	48%	54%	39%	25%
May	6.8%	3.9%	2.8%	3.4%	47%	53%	41%	24%
June	6.5%	3.3%	3.2%	2.6%	50%	54%	44%	28%
Panel B: Week Day								
Monday	7.0%	4.4%	2.5%	4.0%				
Tuesday	5.9%	3.8%	2.0%	3.5%				
Wednesday	5.5%	3.6%	1.9%	3.2%				
Thursday	5.6%	3.6%	2.0%	3.2%				
Friday	6.3%	4.0%	2.3%	3.4 %				

This table contains estimates of absence prevalence and intensity for different types of absence broken out by school month and weekday. Absent, Truant, Excused, and Out Sick are the percent of students with those categories of absence respectively. Truant and Absences in Spell are the percentage of absences that are part of a multi-day spell and Truant to Truant and Excused To Excused transition probabilities are the estimated own transition probabilities for absences. All Students are Freshmen.

Table 1.7: Outcomes by Days Missed and Attendance Type

Days Missed	Continues to 10th	Dropout	GPA under 2.0	ELA Score	Repeat Math	Pass English	Pass Math
Panel A: All Students							
0-2	87%	3.20%	12%	51	4.80%	89%	91%
3-7	89%	5.30%	21%	49	7.20%	88%	88%
8-12	85%	7.10%	30%	47	9.50%	85%	85%
13-17	80%	9.10%	28%	45	11.0%	83%	81%
18-22	72%	13.9%	51%	43	14.0%	78%	79%
23-27	69%	15.7%	57%	40	15.0%	73%	70%
Panel B: Students with Fewer than 3 Truancies							
0-2	87%	3.2%	12%				
3-7	90%	4.4%	19%				
8-12	87%	4.8%	22%				
13-17	86%	5.1%	25%				
18-22	81%	8.7%	30%				
23-27	78%	8.0%	31%				
Panel C: Students with No Truancies							
0-2	88%	2.8%					
3-7	92%	2.4%					
8-12	90%	2.8%					
13-17	89%	3.0%					
18-22	86%	4.2%					
23-27	82%	5.6%					

Student outcomes are broken up by number of absences and number of truant absences. Panel A is the average outcome by number of absences for all students, panel B is only students who are not chronic truants, and panel C is only students with no truancies. English pass and math pass is the percentage of students in each absence group who pass their exit exam on the first try. All students are freshmen.

Table 1.8: Continuation in Traditional School by Chronic Absence and Truancy Status and Grade Transition

	9th to 10th	10th to 11th	11th to 12th
No Chronic Absence or Truancy	88.1%	89.6%	93.0%
Chronic Absence Only	76.6%	80.8%	91.0%
Chronic Truant Only	75.1%	71.5%	75.9%
Both Chronic Absence and Truancy	56.4%	53.8%	59.0%

The table includes the percentage of students who continue in their traditional public school when changing grades based on four categories of attendance behavior. Students have no chronic absence or truancy if they miss less than 10% of the year and have fewer than 3 truancies. Absence only is students who miss 10% or more but fewer than 3 of those absences are truant, truant only is students who are absent less than 10% of the year but have 3 or more truancies, and both is students who miss 10% or more of the year and at least 3 of those absences are truancies.

Table 1.9: Student Outcomes by September of Freshman Year Absences and Truancies

	Continues to 10th	Trad. Grad.	Drop.	GPA	GPA <2.0	ELA Score	Math Score	Repeat Math	Never Pass	Pass English	Pass Math
					Panel A: Number of Total Absences						
0	85%	69%	8%	2.6	24%	.12	.17	8%	3.4%	85%	85%
1	81%	60%	11%	1.7	35%	-.01	-.02	10%	5.2%	81%	81%
2	76%	51%	14%	1.3	44%	-.12	-.11	11%	6.7%	80%	77%
3	67%	42%	15%	1.8	55%	-.26	-.23	13%	9.6%	74%	74%
4	64%	33%	23%	1.6	62%	-.43	-.36	13%	12.8%	68%	68%
5	60%	29%	21%	1.5	64%	-.45	-.34	11%	14.8%	67%	66%
					Panel B: Number of Truancies						
0	84%	67%	9%	2.6	27%	.09	.129	8%	3.80%	84%	85%
1	72%	37%	21%	1.7	60%	-.50	-.39	15%	10.9%	68%	67%
2	61%	24%	24%	1.3	73%	-.75	-.52	14%	18.2%	63%	60%
3	53%	18%	24%	1.9	79%	-.82	-.61	12%	27.2%	54%	47%
4	52%	18%	30%	.98	83%	-.84	-.56	16%	15.7%	56%	56%
5	52%	9%	40%	.93	84%	-.91	-.61	10%	30.4%	44%	41%
					Panel C: Number of Excused Absences						
0	84%	66%	9%	2.6	27%	.07	.13	8%	4.2%	84%	84%
1	81%	61%	10%	2.4	33%	.02	.01	9%	4.7%	83%	82%
2	77%	53%	12%	2.2	41%	-.05	-.06	10%	5.7%	82%	78%
3	70%	48%	14%	1.9	49%	-.09	-.16	13%	6.6%	79%	79%
4	66%	41%	19%	1.8	54%	-.27	-.22	12%	10.3%	74%	73%
5	67%	38%	15%	1.7	56%	-.31	-.33	10%	11.7%	72%	72%
					Panel D: Number of Truancies for Students with No Truancies						
0	85%	69%	8%	2.6	24%	.12	.18	7%	3.4%	85%	86%
1	83%	63%	9%	2.5	30%	.08	.04	9%	4.0%	84%	84%
2	80%	58%	10%	2.3	36%	.01	-.01	10%	4.5%	83%	80%
3	73%	52%	11%	2.1	44%	-.02	-.09	12%	5.4%	82%	81%
4	70%	45%	19%	2.0	48%	-.20	-.20	11%	8.7%	76%	76%
5	70%	43%	12%	1.9	51%	-.23	-.28	10%	9.7%	74%	77%

This table includes averages for 11 outcome measures of student achievement broken out by the number of absences a student has during September of her freshman year by absence type. Panel A is the number of total absences, panel B is the number or truant absences, Panel C is the number of excused absences, and Panel D is the number of excused absences and includes only students with no truancies whereas panels A - C contain all students. ELA and Math score are standard normal test scores and the units are standard deviation. Test scores are 9th grade standardized tests, and never pass, pass English, and pass Math are high school exit exam pass rates.

1.8 Figures

Figure 1.1: Demographic and Truant Differences Across Schools

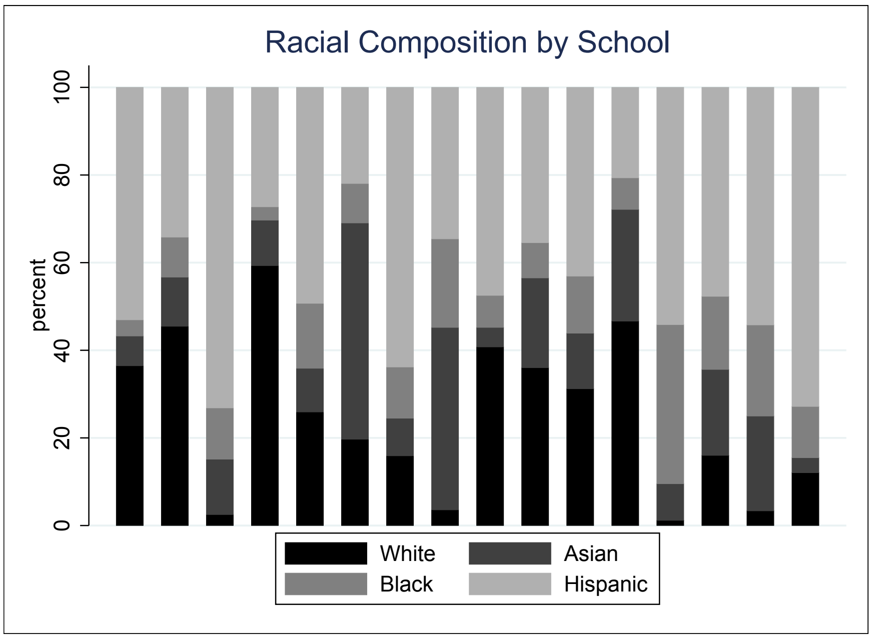
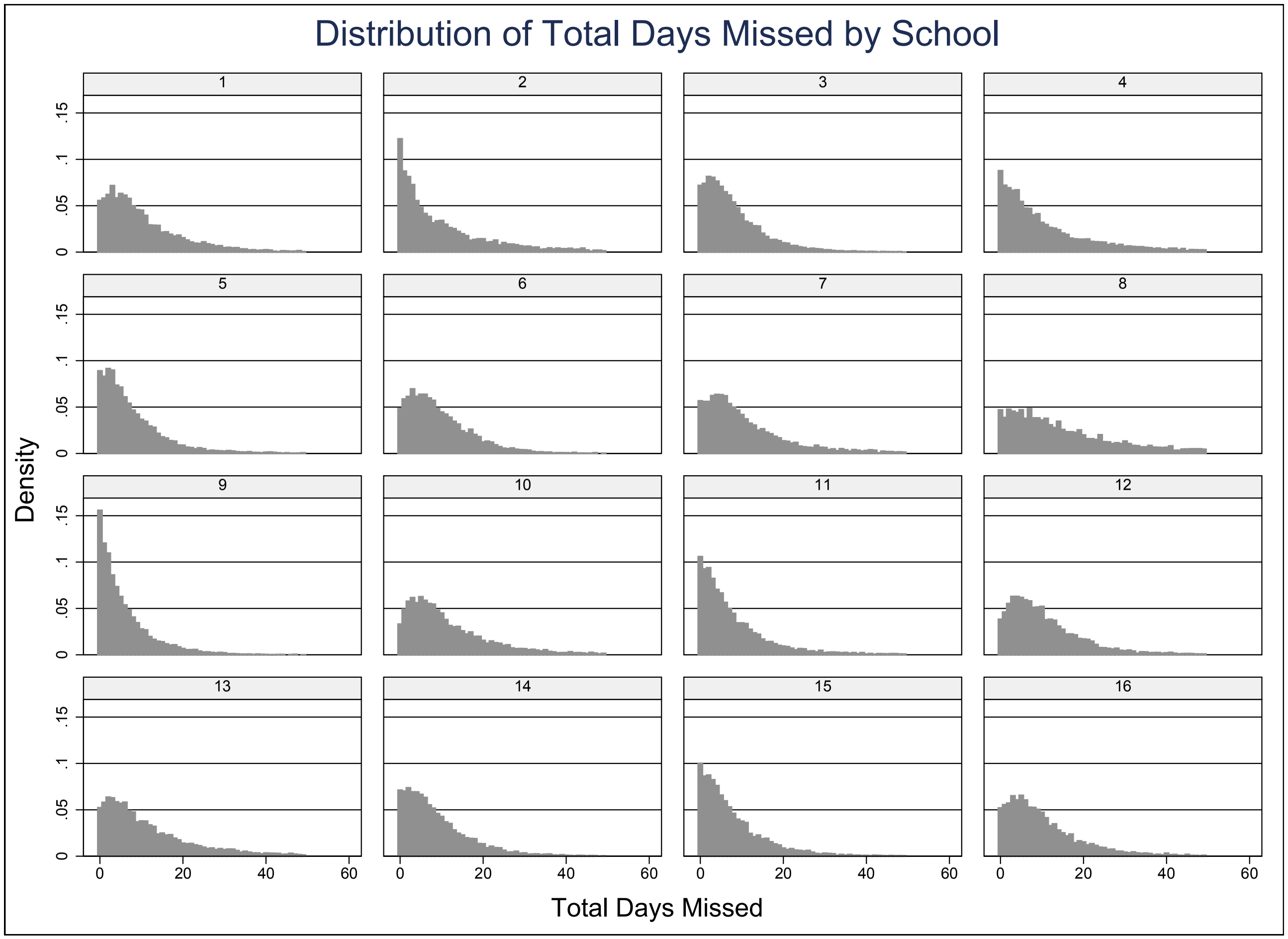


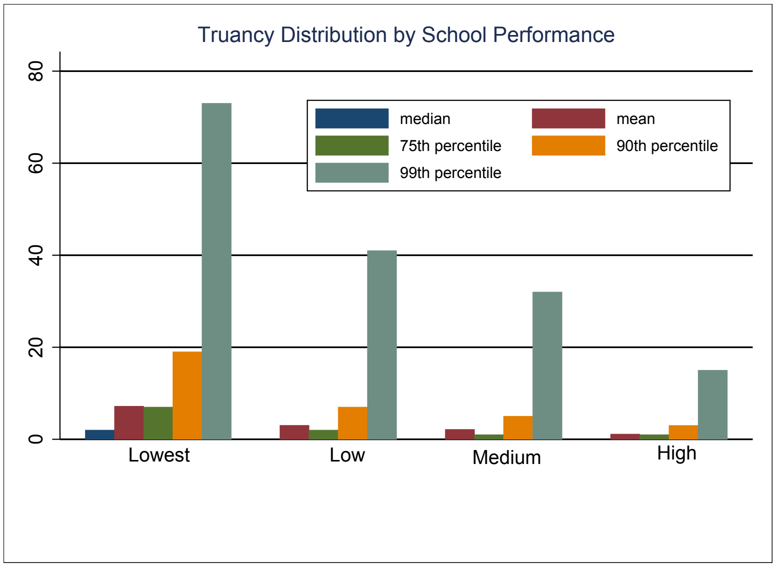
Figure is based on author's calculations from selected empirical sample of entering 9th graders. Each column is an individual school and racial categories are mutually exclusive.

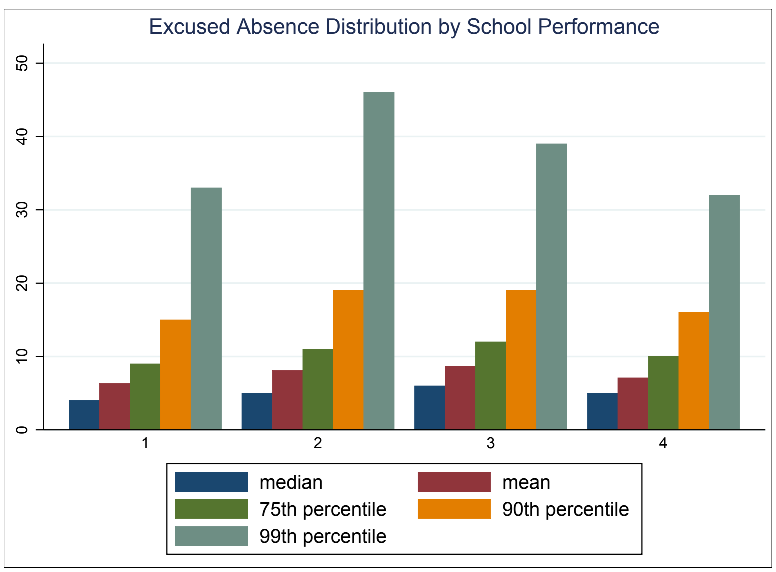
Figure 1.2: Distribution of Total Days Missed by School



Each histogram is an individual school and the height of each bar of the histogram is the number of students in the school who missed that many days of high school during freshman year.

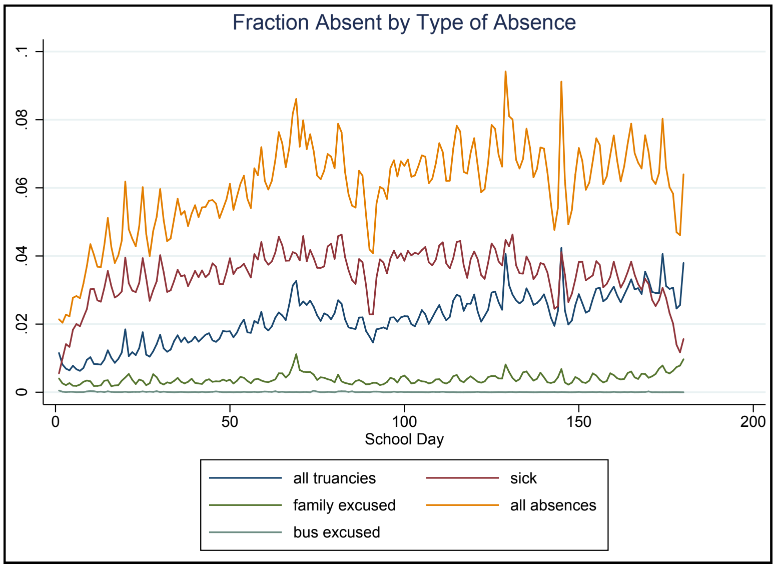
Figure 1.3: Truancy and Excused Absence Distribution by School Performance

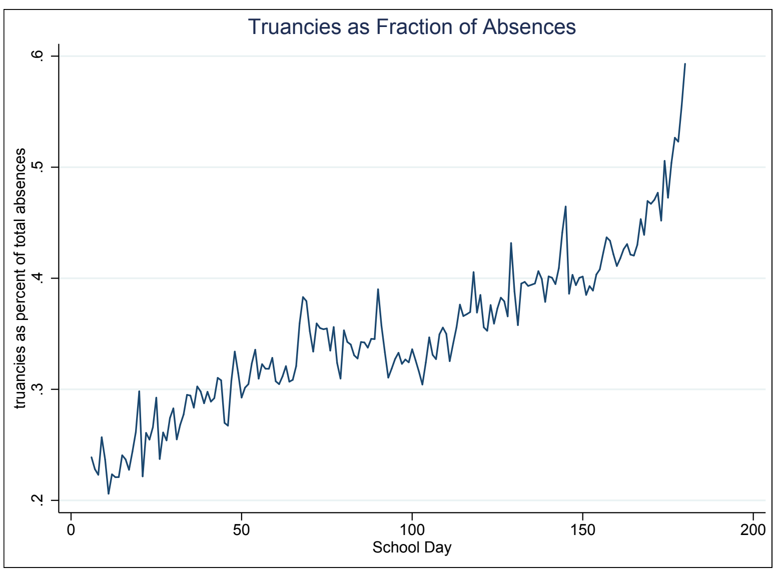




Bars are the mean, median, and percentiles of the truancy (top) and excused absence (bottom) distributions by school performance group ranging from lowest performing (1) to high performing (4). Performance quartiles are based on test scores, graduation rates, and grade point average.

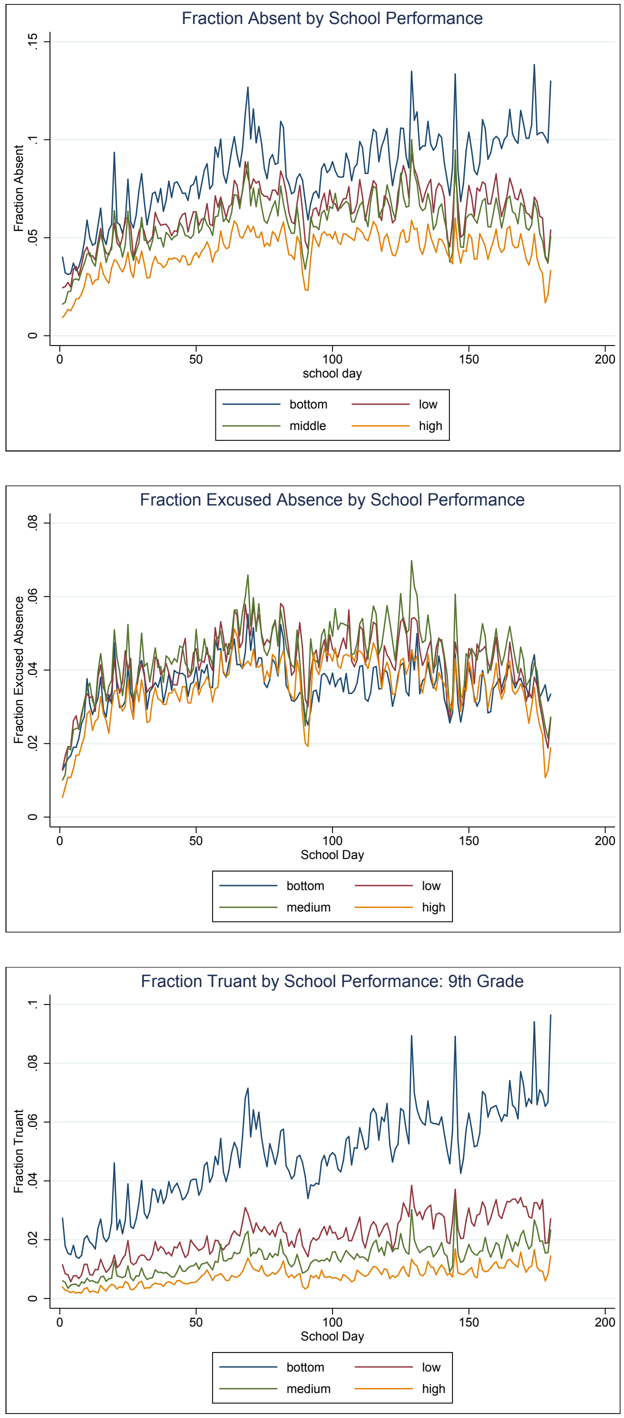
Figure 1.4: Trends in Absence Type





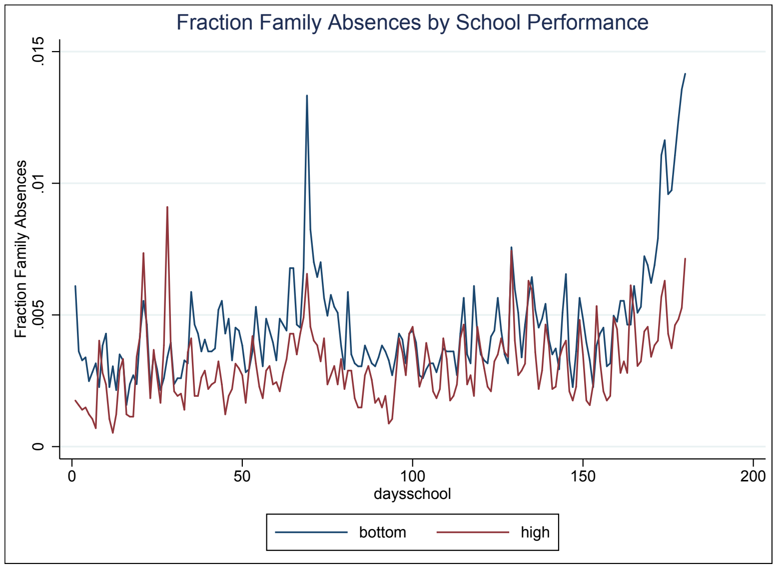
Figures contain average daily attendance trends during freshman year by absence type. The top panel shows the trends across the year for all absences, sick leave, family leave, bus exceptions, and truant absences. The bottom panel is the fraction of absences that are truant.

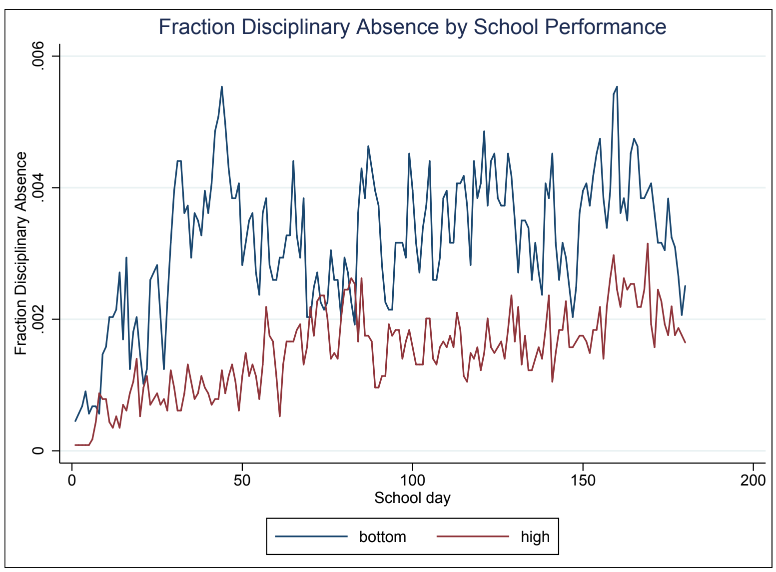
Figure 1.5: School Year Absence Trends by School Performance



Figures contain average daily attendance trends during freshman year by school performance quantile for total absences (top), excused absences (middle), and truant absences(bottom). Each data point is a school day.

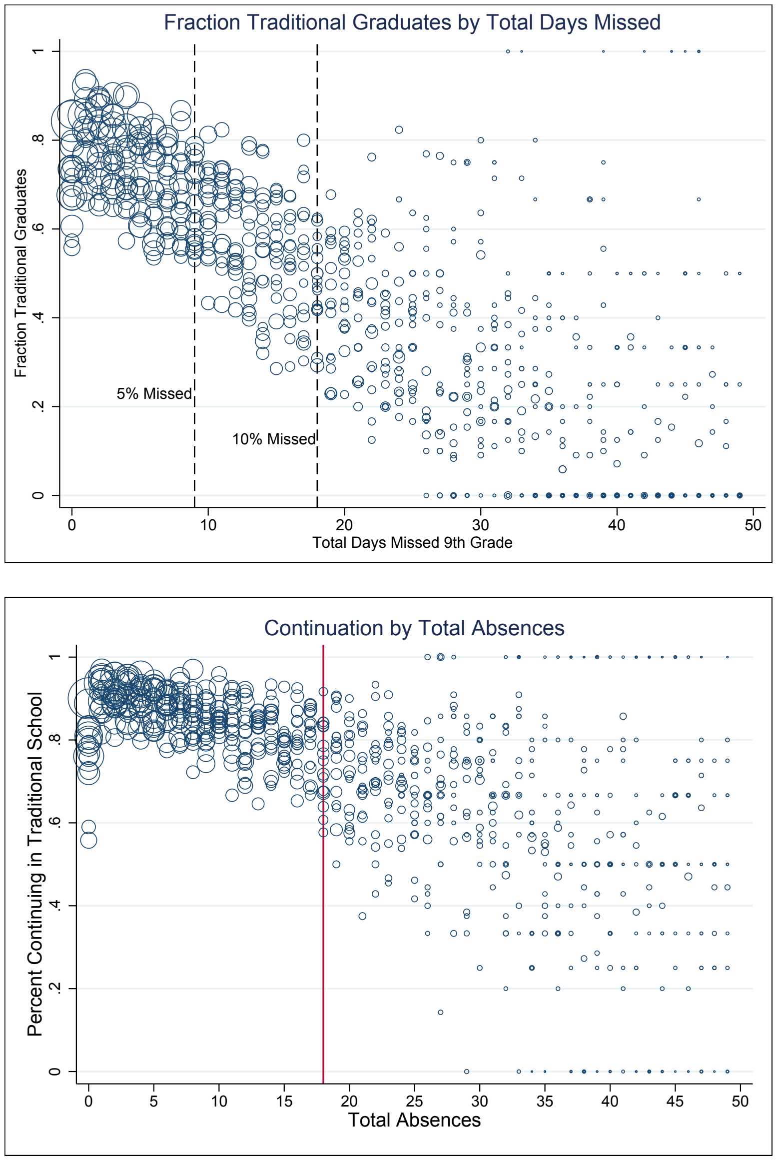
Figure 1.6: School Year Absence Trends by School Performance





Figures contain average daily attendance trends during freshman year by school performance quantile for disciplinary absences (top and family emergency absences (bottom). Each data point is a school day.

Figure 1.7: Continuation to 10th Grade and Graduation by Total Absences

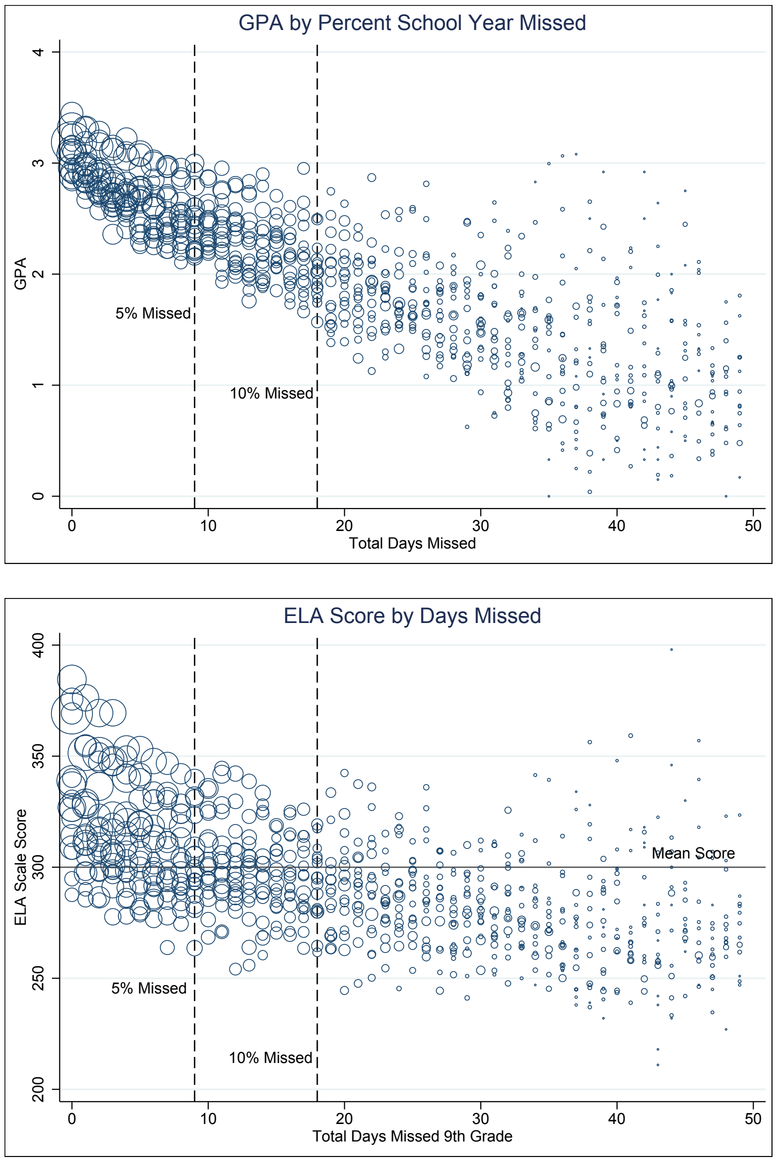


Each circle is a school by number of days missed group and are scaled in size by the number of students in the school by days missed cell. The y axis of the top panel is the fraction of students in each cell that graduate from a traditional high school and the y axis on the bottom panel is the percent of students continuing in traditional school from 9th to 10th grade. Reference lines make the points in which a student would have missed 5% and 10% of the school year.

Figure 1.8: School Year Absence Type Trends by 10th Grade Enrollment

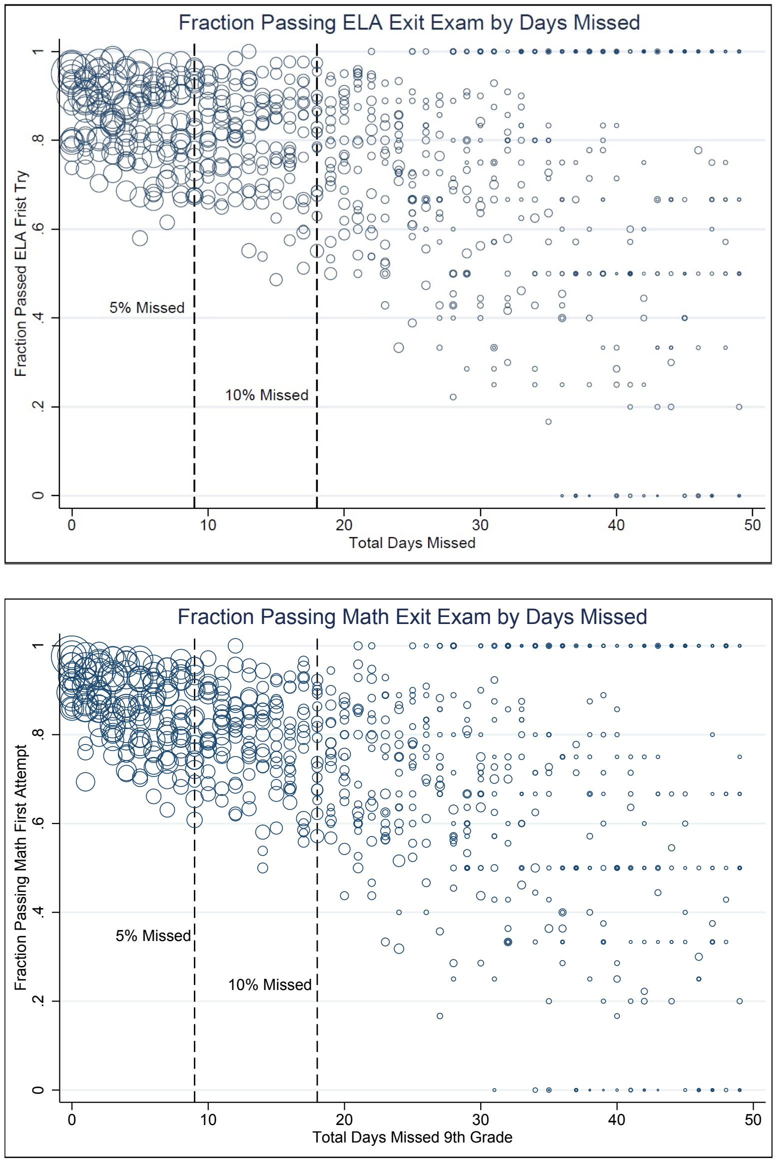
Figures contain average daily attendance trends during freshman year by 10th grade enrollment for sick absences (top), truant absences (middle), and disciplinary absences(bottom). Each data point is a school day. Top and bottom panels only include students who continue in some type of school, the middle panel includes all students.

Figure 1.9: GPA and ELA Score by Days Missed



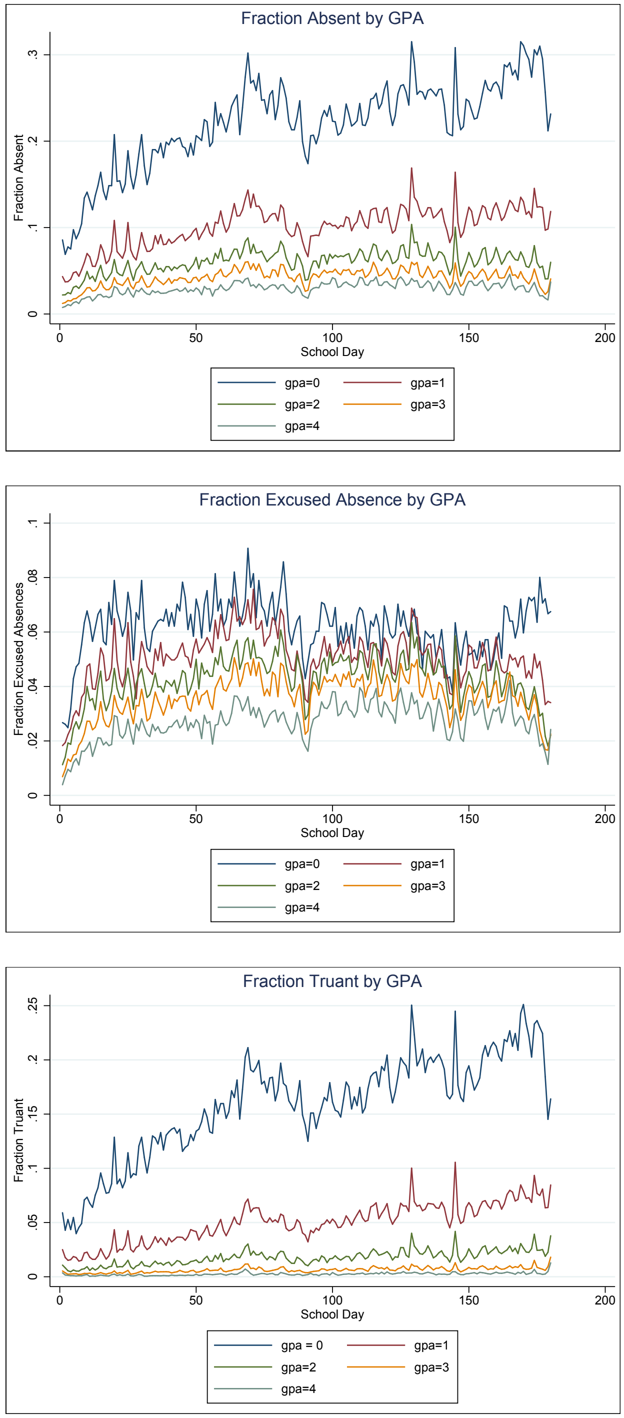
Scatterplots contain means of GPA (top) and 9th grade ELA score (bottom) exit examinations. Each circle is a school by number of days missed group and are scaled in size by the number of students in the school by days missed cell. Reference lines make the points in which a student would have missed 5% and 10% of the school year.

Figure 1.10: Exit Exam Pass Rates by Days Missed



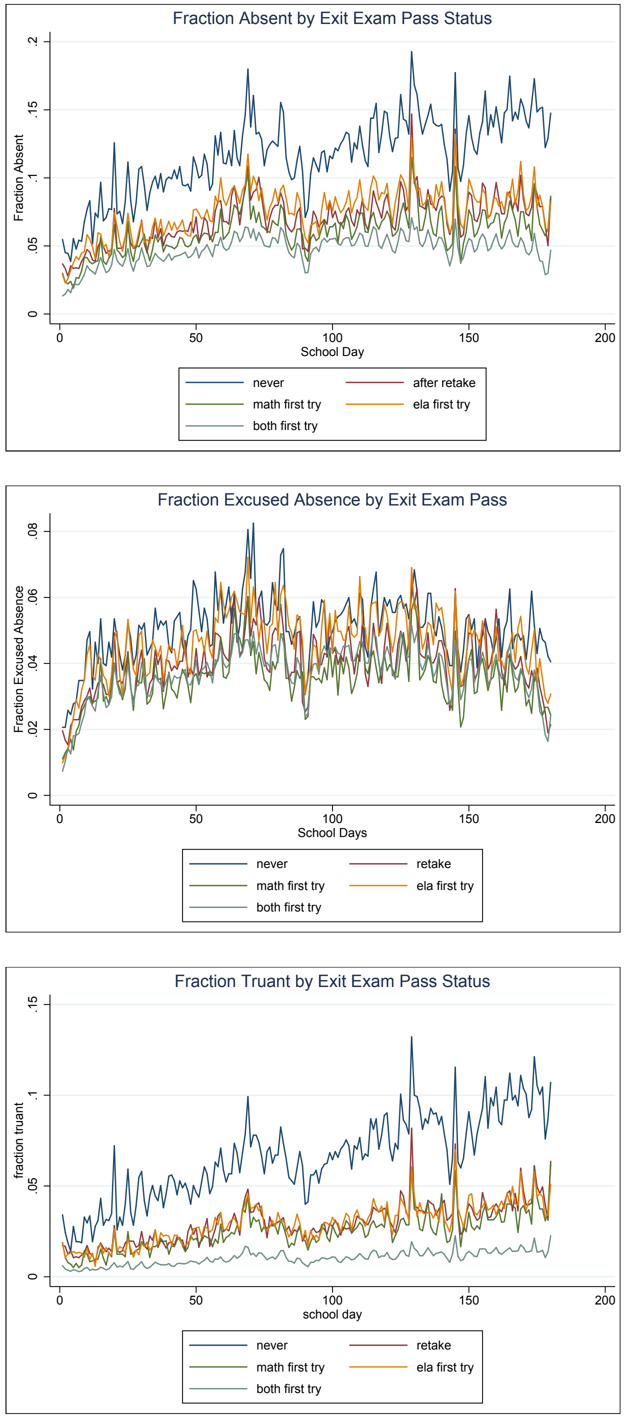
Scatterplots contain means of first time passage rates for the ELA (top) and Math (bottom) exit examinations. Each circle is a school by number of days missed group and are scaled in size by the number of students in the school by days missed cell. Reference lines make the points in which a student would have missed 5% and 10% of the school year.

Figure 1.11: School Year Absence Trends by GPA



Figures contain average daily attendance trends during freshman year by weighted GPA for total absences (top), excused absences (middle), and truant absences(bottom). Each data point is a school day.

Figure 1.12: School Year Absence Trends by Exit Exam Pass Type



Figures contain average daily attendance trends during freshman year by which exit examinations the student passed on the first try if at all for total absences (top), excused absences (middle), and truant absences(bottom). Each data point is a school day.

CHAPTER 2

THE DYNAMICS OF HIGH SCHOOL DROPOUT

2.1 Introduction

A large body of evidence finds that there are high lifetime returns to graduating from high school compared to dropping out or completing a high school equivalency exam (GED) (Angrist & Krueger, 1991; Cameron & Heckman, 1993; Card, 2001). Studies have found that students who are induced to stay in school longer because of mandatory schooling laws have higher income over their lifetimes (Rouse, 2008, Oreopoulos, 2007), are less likely to report being in poor health (Muenning, 2008; Link & Phelan, 2005; Wong et. al., 2002; Lleras-Muney, 2004), are less likely to be unemployed, and less likely to report being unhappy. Higher graduation rates are associated with declines in incarceration rates (Lochner & Moretti, 2004) and school attendance reduces juvenile property crime (Jacob & Lefgren, 2003). Additionally, many high school dropouts regret the decision to drop out later in life. Bridgeland et. al (2006) conducted a survey of young adult dropouts and found that 81% of students said that graduating high school was important for success and that 74% would have stayed in high school if they could do it over.

Despite the apparent benefits of high school and dropouts' own beliefs about the importance of graduation, dropout rates remain persistently high with estimates varying from 20-30% of high school students.[1] Even the Department of Education's high end official estimate of 82.3% (NCES, 2015) is still far below

[1] There is extensive debate on how to measure dropout rates. See Belfield & Levin (2008) for an overview of the methodology.

the OECD average and the US graduation rate places below that of Slovenia, Portugal, Hungary, and the Slovak Republic (Murnane, 2013; OECD, 2013). Additionally, this high average official graduation rate is not representative of all students. Dropouts are disproportionately likely to be poor, Black, Hispanic, and to live in a few districts, mostly large urban school districts, with very high dropout rates (NCES, 2015). All five of the US's largest districts have dropout rates over 30%.

Labor economic studies of dropout and school attainment that seek to explain why graduation rates are low in the face of high measured returns to high school, are generally implicitly or explicitly based in the Becker human capital accumulation framework (Becker, 1962). In this framework, students are rational forward looking economic actors and choose how much schooling they plan to attend based on the expected costs and benefits of that schooling. Students will attend until the marginal cost of doing so is equal to the marginal benefit and drop out if the benefits to graduation are not high enough to justify the cost. However, dropout is not a single event or decision, rather it is the outcome of a series of small decisions over everything from finishing an assignment to catching the right bus. It is unlikely that students have perfect knowledge of the costs and benefits of schooling or the education production function itself. Furthermore, the traditional Becker framework predicts that students should attend school all at once, not alternate between periods in and out of school. Increasingly severe truancy and academic problems before eventual dropout is contrary to the predictions of a traditional Becker model. However, it is exactly these patterns that are documented in the literature as precursors to dropout (Christenson et al., 2012; Rhumberger, 2004; Bridgeland et. al., 2006).

Understanding the process by which students make education input decisions and accumulate schooling is critical to design effective dropout and truancy prevention policies. In the case in which students are not making perfectly rational or informed education investment decisions, either because they are shortsighted or do not understand the education production function, policies that focus on lowering the cost of schooling for students may be more effective than polices that stress the benefits. The set of policies that will successfully reduce high school dropout rates depends critically on the decision model that high school students follow and understanding the incentive students face.

This paper proposes a basic theoretical model of education investment to motivate empirical results on how students respond to changes in the cost of schooling, and how those cost changes can lead students to have snowballing truancies and drop out of high school. The goal is to explain the persistence of high school dropout despite the high measured returns to graduation and remorse felt by dropouts themselves. I develop a model based in the Becker human capital accumulation framework in which high school education is modeled as a customer-input technology and students do not have full knowledge of the production function, specifically the relationship between present inputs and future costs. Students are both the consumers and producers of education and their choices over inputs affect not only the amount of schooling they receive (high school diploma, GED, dropout) but also the value of that schooling in terms of diploma quality and human capital accumulation.

I model two types of decisions that students make; the daily decision whether to attend high school or be truant each day, and the decision after the end of the school year whether to return for the following year. In the model, students know that school is valuable in the future and that missing a day decreases both the value of their credentials and the probability of graduating, but they are ignorant of the full dynamics of the education production function. The key component of this model is that it allows students who ex-ante plan to graduate to become off track and not complete high school despite ex-ante plans to do so making dropout a gradual process of increasing truancy. It predicts that students with a low cost threshold for school will have more frequent absences over time and that spells of absence will be longer over time. This model is in line with research on educational "engagement," which views dropout as part of a process in which students gradually disengage from school.[2]

To analyze the dynamic relationship between truancy, schooling cost, and dropout more generally, I employ a novel and unique data set from a large urban school district (LUSD) that contains over 25 million observations of daily level attendance information for over 37,000 students. On each day of school, while the student is enrolled in their initial high school, I observe if she is in school, and if she is not in school, whether it is an excused or truant absence. I also observe when students transfer within or out of the district, if and when they graduate, and what school or program they graduate from. In order to evaluate how students respond to cost changes, and how those cost changes affect attendance decisions and potentially snowball over time, I exploit an opportunity cost increase faced by students when they turn 16. At 16 students' are allowed to legally work more jobs and longer hours. They are also no longer subject to juvenile court truancy enforcement even through the mandatory schooling age is 18. Different students will face the cost shock at different points in their academic careers based on their birth date. This provides plausi-

[2]See Christenson et al. (2012) and Rhumberger (2004) for overviews

bly exogenous variation in the timing of a large increase in the opportunity cost of schooling.

Consistent with the motivating model, I estimate the effect of turning 16 on a student's decision to attend school each day and the effect of those truancy decisions on high school graduation. A key prediction of the model is that truancy should not necessarily increase immediately following the increase in opportunity cost at age 16. For daily attendance I use a double fixed effect framework controlling for a student and day fixed effect to account for student fixed ability and for any characteristics of the school day that are likely to affect whether or not students attend on that day. The time a student is 16 enters as a quadratic spline and the length of time a student has been 16 is allowed to have a non-linear relationship with attendance. Students who are 16 longer are more likely to be truant and are truant more often. I present both OLS and IV estimates for the effect of truancy on dropout. The IV specification instruments total yearly truancy with a quadratic spline of how long the student is 16 similar to the daily level results.

I find that an increase in the opportunity cost of schooling increases the probability that a student will be truant, and that the probability becomes larger the longer the student is exposed to the higher opportunity cost. Students who have already had attendance problems before turning 16 face even greater increases in their truancy than other students. A district-labeled "chronic truant" who spends three additional months above the age of 16 during her sophomore year is up to 69% more likely to be truant each day going forwards. Additionally, students with more truancies are less likely to graduate, and if they do graduate, are more likely to do so with lower quality alternative degrees. A student with 3 truancies is 10% less likely to continue in her traditional school the following year. Turning 16 does not have the same effect on excused absences which remain unchanged when a student turns 16. There is also no long term dynamic effect from turning 15. The results support existing empirical evidence that the most effective dropout prevention policies for students, once they reach high school, are those that decrease the daily cost of schooling. These policies include better truancy enforcement (Lemieux et al, 2011) or intensive tutoring and mentorship programs that increases students' desire to be in school and their ability to complete coursework (Bell et. al., 2015; Oreopoulos et al, 2014; DeSocio et al, 2007). The results also demonstrate on the importance of understanding the long term dynamic effects of education policies on truancy. A traditional regression discontinuity framework would have shown no effect of turning 16 because students develop truancy problems gradually as they face higher costs of school. It is likely that other education policies, especially those that permanently increase or decrease the cost of attending school, may have similar long run effects on student truancy

After discussing the motivating dropout prevention and theory literature, the paper proceeds as follows. First, I discuss the novel data set of daily level student attendance information linked with education outcomes that I employ in this paper. I then develop a theoretical model of high school dropout to motivate my empirical econometric estimation. The model is based in the Becker framework and incorporates aspects of education engagement theory. This model predicts that students may ex-ante plan to graduate, appear to have the ability to graduate, but then still become "off track" by missing more days of school than planned and, most importantly, that an opportunity cost increase can have a long term effect on truancy rather than an immediate effect. I then describe my two-part empirical strategy in which I estimate the daily effect of the increase in opportunity cost at age 16 on daily truancy decisions using a double fixed effect model with plausibly exogenous timing in a cost increase at a student's 16th birthday, and the between year effect of accumulated truancies on continuation in traditional school at the end of each grade in both an OLS and IV framework using the time a student was subject to the increased daily opportunity cost as the instrument. I then conclude with the empirical results and a discussion of policy implications.

2.2 Dropout Theory Literature

Because of the importance of dropout prevention, many researchers have studied the causes, consequences, and correlates of dropout. Some prominent interventions that have been examined in prior research include parent involvement, drug use, school lunch programs, cash transfers, scholarships, student-teacher relationships, and pregnancy (Walker et al 1998; Filmer & Schady, 2011; Marshall, 2011; Roebuck et al, 2004; De Witte & Csillag, 2012; Lovenheim, Reback & Wedenoja, 2015). However, there is a dearth of causal evidence because many prevention programs are not implemented in a testable way. The studies of the demographic, institutional, and community predictors of dropout are extensive and include race, ethnicity, gender, parents' education, income, neighborhood crime, class size, and many others.[3]

The most common model that labor economists use to measure the causes and effects of high school dropout, and school attainment more generally, is

[3] See Rhumberger (2001) for a detailed overview.

the Becker human capital accumulation model (Becker, 1964; Mincer, 1958). In this model, students are rational forward looking economic agents and decide on their level of schooling in order to maximize their lifetime utility. Students will attend school until the marginal benefits of continued attendance are outweighed by the marginal costs. The model has been expanded to take into account that students do not make all of their schooling decisions at once, rather that they make sequential decisions and decide before beginning an additional year of schooling whether to start that year taking into account the probability of not finishing the year (Cameron & Heckman, 1998,2001; Cunha&Heckman 2007, 2008, 2010; Eckstein&Wolpin, 1999).

The idea of high school dropout as the outcome of a series of small decisions and events, such as truancy, is often referred to as "disengagement" from school (Christenson et. al., 2012; Rhumberger 2004 for overviews). While social scientists disagree about the exact mechanisms through which disengagement develops, it can be observed as an increase in truancies and absences, failing grades, behavior problems, and a lack of relationships with teachers or peers. In the general framework of engagement and disengagement theory, engagement is an unobservable quality that students have related to their ability, attitude, upbringing, and other personal factors and how those traits interact with the school and community around them. Engagement is not perfectly observable but students who are more engaged are more likely to stay in school. Disengagement could happen for a variety of reasons. Students could be frustrated after school failures leading to the development of behavioral and other problems and eventually the decision to leave school or the inability to graduate (Finn 1989). Students may also not feel that they belong at school because they are not part of school sponsored activities (Wehlage et all 1989). In addition to push factors discussed above, students could face pull factors from outside the school such as job opportunities, family obligations, or friendships with peers who are not attending school.

Dropout is only one of the possible consequences of a disengagement process manifested through truancy and failing grades. Transferring to another school, especially within the same district, is a less severe form of disengagement than completely dropping out (Lee & Burkham, 1992; Rhumberger & Larson, 1998; Rhumberger 2003). Some students who disengage from school may leave their traditional public high school for an alternative school for similar reasons and through a similar process as those that lead other students to drop out of school entirely.

2.3 Data and Descriptive Statistics

Data for this project come from a Large Urban School District (LUSD) and contain daily attendance records, demographic information, entrance, exit, and transfer information, and test scores for five cohorts of freshman in traditional neighborhood public schools. The dataset follows these students until they drop out, graduate, or otherwise leave the district. The data set is at the student by year and student by day level. The data include over 25 million observations for 37,000 students and are unique in that they include students' detailed attendance by date for every day of their high school careers and links that information to their educational outcomes.

2.3.1 Student Record Files

Demographic data for the students come from LUSD's student entrance file. The entrance file includes all students who enter into a traditional public high school between the years 2005 and 2009. The student file includes the entrance school, school year, student demographics, and birth date and is augmented by the student exit file that includes the last school that the student attended in the district and the date and reason for a student's exit: graduation, dropout, or transfer. The exit file contains detailed administrative exit information generally unavailable in survey datasets. It distinguishes between types of transfers (in distinct, public in state, private in state, out of state, out of country) and types of diplomas that students earned. Diploma types include a high school diploma from a traditional school, a high school diploma from an alternative school,[4] an adult education diploma, a GED, or no diploma.

The data also contain students' high school exit exam scores, yearly weighted grade point averages,[5] and accountability test scores in math and language arts. Students in the same grade all take the same language arts test[6] but the math test is for the specific math class in which the student is enrolled.[7] The high school exit examination is required for graduation and generally first attempted and usually passed in the spring of tenth grade; 80% of 10th graders

[4]I define alternative schools as public high schools within LUSD that waive LUSD's standard high school requirements. These schools may have non-standard school days, online classes, fewer credits required for graduation, and may waive the high school exit exam and other required benchmarks. Gifted schools and schools with special programs that do not waive requirements, such as arts or technology high schools, are not considered alternative schools.

[5]The weighted GPA is the one the district uses to determine if a student has passed the GPA threshold of 2.0 to graduate.

[6]It is extremely uncommon for a student to repeat a grade's language arts test; fewer than 1% take the same test two years in a row.

[7]99% of students start in Algebra 1 or the equivalent and only 9% of students repeat the same level of math the following year.

pass the exams before 11th grade, although students may continue to re-take the test until they pass.

2.3.2 Sample Selection

In this paper I focus on a sample of "traditional" public high school students rather than restricting my analysis to "high risk" students. While many of the risk factors associated with dropout are apparent in middle and early high school, there are many students who appear capable of graduating upon entry into high school and yet do not graduate. These are potentially students who do not get dropout prevention services because they are not labeled high risk and by the time that they start to exhibit problems are considered too old to help. School districts have limited resources and services are often directed towards younger students where they are believed to be more cost effective and where they will have a longer term to be successful.

For this paper, I define traditional students as those who begin high school in 9th grade at their neighborhood school, are between the ages of 13.75 and 15.5,[8] and are not in non-diploma bound special education programs.[9] The resultant sample of students is slightly higher achieving than the average 9th grade student in the district in terms of GPA, graduation rate, and test scores. Most older students who have been held back in the past are excluded from the sample as

[8]State law requires students be at least 5 years old by September 1st when they begin kindergarten in public schools, which equates to 14 years old in 9th grade and students must attend school if they turned 6 years old by September 1st equating to 15 years old in 9th grade. Private schools and other states do not have the same rules so students who began kindergarten elsewhere may be younger than 14 or older than 15.

[9]Students with Individual Education Plans (IEPs), who are enrolled as diploma bound students are included in the sample and can't be distinguished from other students.

are students who begin high school already in alternative programs.[10] All students in the sample take the grade level language arts exam in their first year and take at least eighth-ninth transitional math.[11] The sample also excludes students who begin ninth grade late, after September, or who leave the district before the end of September of their freshman year.

The specific school a student attends is based on residential district zoning and as a result, the demographics of each school in the sample are different and reflect the neighborhood in which the school is located. There are 16 high school campuses in the sample with graduation cohorts ranging from 200-600 students. The district is majority minority: 40% Hispanic, 10% Black, 27% White, and 20% Asian. However, due to the neighborhood catchment areas of the schools, that the racial composition of individual schools varies greatly. Table 1 shows the difference in student outcomes across schools in the district. Schools vary greatly in the outcomes of their students in terms of transfer, graduation, and dropout rates. The percentage of eventual district graduates ranges from 48% to 79%.[12]

2.3.3 Attendance Data and Description

The primary data I use is LUSD's daily final attendance code dataset. This dataset contains extremely detailed information on the daily behavior of high school students. The dataset contains one attendance code per student per day and distinguishes between excused absences and truancy. The primary focus of

[10] I have not included official district metrics or differences from my measures in order to keep the district anonymous.

[11] Over 90% take Algebra 1 or the equivalent.

[12] Eventual transfers are included in the denominator.

this paper is truancy rather than excused or all absences. I use the legal definition of truancy: a student is truant if she is not in school on a day in which she is legally required to be there and does not have a legally valid excuse for the absence. Students are required to be in school if they are below the mandatory schooling age of 18 and do not have a high school diploma or GED, have not transferred officially to a school in another district, are not officially enrolled in a home school program, and are not incarcerated or institutionalized. Students also do not have unlimited excused absences. Beyond three days of illness students are required to have a doctor's permission and documentation of the severity of the illness and many ill students still 'attend' school remotely which is coded in the attendance file. Furthermore, parents can only excuse students for illnesses, bereavement, and religious holidays all of which have day caps and can require documentation. A parent's permission is not enough to justify a student's absence and absences beyond these limits, even if excused by the parents, are still truant absences.[13]

Students who do attend school fall into two categories: present and on-time, referred to as "present" going forward and present and late, referred to as "tardy." On average, 85% of students attending traditional public high schools are present and on time, 6% are tardy, 4.2% have excused absences and 2% are truant. The percentage of students that are truant on an average day is very low and does not remain constant across grades. The rate is highest freshman year (2.15%) and is lowest senior year at 1% and increases over the course of the school year from 0.82% in September to 2.69% in June. Truancy is widespread among students but not universal: 72% of students will be truant at some point

[13]The final attendance code data set distinguishes between "Truant confirmed by parent" and "unexcused" or "unverified" absence all of which fit the legal definition of truancy. The majority are "Truant confirmed by parent."

during their high school career and 28% will never be.[14]

Truancy is low overall but are concentrated in a smaller number of chronically truant students and students in low performing high schools. Close to one third (28%) of freshman will be classified as a chronic truant, meaning that they have missed over 3 days so far that year. Only 68% of ninth grade chronic truants will continue in their neighborhood public school compared to 88% of other students. There is also substantial variation across demographic groups in both truancy and graduation as seen in Table 2. Black and Hispanic students have higher levels of truancy and are less likely to graduate from traditional programs. Nearly half of Black and Hispanic men will not graduate from a traditional school compared to less than 20% of Asian women and less than 30% of White students.

A student is a "dropout" if she leaves LUSD without a diploma and has not transferred to another school district, hence students can only drop out between school years. The intuition behind this definition is that students choose to begin each school year and then, during that school year, decide each day whether to attend or not. Even if a student drops out in the middle of the year and no longer comes to school, that student still legally has the option (and requirement) to return the next year and is not counted as a "dropout" until she fails to re-enroll or transfer the following year. While most students who have extensive truancies or disappear for months at the end of the school year do not return the following year, some do, and should not be falsely labeled as dropouts. I focus on the attendance behavior students exhibit while enrolled in traditional public school, and how that behavior affects transfer to alternative

[14]See Wedenoja (2017) for detailed descriptions of attendance and truancy patterns and their correlations with student outcomes.

high schools between the school year in addition to how that behavior affects dropout.

2.4 A Model of High School Attendance and Dropout

2.4.1 Motivation and Overview

The goal of this model is to provide a motivating framework for the empirical results by incorporating daily level attendance choices into a Becker Human Capital framework. There are four relevant predictions from this model that are consistent with the empirical results. First, students in this model will not attend school all at once. Some students may have alternating periods of truancy and attendance even missing large sections of a school year and returning the following year. Second, students who ex-ante plan to graduate and have the ability to do so may still drop out of high school or transfer into an alternative program by becoming off track. Third, problems with truancy increase over time as students face higher costs to schooling from missing earlier days. Fourth, students who face a large permanent cost increase may not immediately change their behavior, rather the longer the student is exposed to the cost shock, the more likely she is to miss school.

Following the engagement and disengagement literature, the model treats dropout as a process rather than an event. The model expands the standard Becker human capital accumulation framework to daily decisions and makes more lenient assumptions about students' control over education outcomes and their knowledge of the education production function. Specifically, each year students make the decision whether or not to begin another year of school based on the expected costs and benefits, then, each day, make the decision to attend school that day or not. They do not have full knowledge of the education production function so they are not entirely aware of how the seemingly small daily decisions they make affect the costs and returns they will face in the future.[15] Neither high school dropout nor the choice to achieve another year of schooling are single decisions as they are often portrayed in traditional Becker style models. Enrolling in another year of school may be a single decision but whether that year is completed is the result of the series of education input choices including attendance, assignment completion, and effort that follow enrollment.

These effort investment choices can be complex. Students, with varying degrees of autonomy from their parents, select the school they will attend, the classes they take, how often to attend, how much time to spend on homework, and whether or not to pay attention while they are in class, all of which contribute to the probability that a student graduates and the market value of the credential she earns. A high school diploma has a number of requirements, in the case of LUSD those requirements include attending eight semesters for 6 periods per day, maintaining a GPA of 2.0 or higher, completing specific required courses, and passing the state's high school exit exam. Students need to manage trade-offs and invest their time and effort accordingly over four years in order to meet all of these requirements and graduate.

The model incorporates insights from the disengagement literature into the Becker model in two ways. First, it expands the choice set that students face from beyond years or levels of schooling to choosing daily inputs of attendance

[15]There is substantial evidence adolescents are less able than adults to consider future consequences and are often hyper-responsive to immediate rewards. See Blakemore & Robbins (2012) and Casey et al. (2008) for overviews of adolescent brains and decision making.

and truancy. Second it expands the possible outcomes that students can face. In addition to dropout, the model incorporates the decisions of students to leave their current school to attend an alternative school or receive a GED as part of the education investment process.

Students' actions determine the future value of the high school credential type, if any, they earn and the expected value of the credential affects whether or not the student believes attempting to earn the credential is worth it. Schooling is a customer-input technology because students are both the consumers and producers of its value. Not all high school credentials have the same value. Students who complete the GED or Adult Education Diploma (a separate degree), or who have some of their requirements waived by attending an alternative program are likely to have lower returns to those credentials than they would to a traditional high school diploma (Cameron & Heckman, 1993). Additionally, students who attend alternative programs with flexible schedules or online leaning may miss out on developing valuable non-cognitive skills through interaction with peers and developing skills to navigate a consistent daily schedule as is generally required in the workforce.

Students make two different decisions: the decision to begin a year of schooling and the decision, after deciding to enroll, to attend school on each day or to be truant. Students make this decision as traditional forward looking agents but as ones who do not have all of the information about the education production function. At the beginning of each year a student has an idea of the benefits of completing that year of schooling and the distribution of daily costs that she will face over the year in order to complete it. Using this information, she will decide whether or not the expected benefit of that year outweighs the expected total cost. She considers how often her predicted daily costs will be so high that she will not attend and how that will affect her human capital accumulation and degree quality. She understands that if she misses too many days of school that she will be unable to complete the year and graduate. If her expected truancy is high enough that she expects not to finish the year, she will choose not to start it. However, while she weighs how truancy will affect her returns to high school, she does not consider the effect of truancy on its cost. Missing more days makes it harder to graduate because she will have learned less. However, each day missed also increases the cost of attending on future days through having to catch up on school assignments, losing non-cognitive skills like the ability to get up in the morning or do schoolwork for a sustained seven to eight hours, and a loss of social connection at school. When deciding each day whether or not to attend, the student does not take into account how the decision will increase the costs she faces in the future. This can lead a student who began a year of school with the intent to finish, and who believed that she had an achievable plan to finish school, to drop out despite those ex-ante plans. After missing a few days, she faces higher than expected costs, which cause her to miss even more days. This may eventually lead her to dropout or transfer to an alternative program that has lower costs due to waived requirements and lower returns because of those waived requirements.

2.4.2 Costs

School is costly to students both directly, through the effort required, and indirectly from the opportunity costs of the activities students give up while in school such as work, time with friends, family responsibilities, or sleep. Students who have negative feelings about school because they have few friends, are tired all the time, have low self esteem, do not understand the material, or who fell physically bad having to sit in a desk all day, face higher daily costs of schooling as well. Costs can be broken into two categories: shared costs and individual costs. Shared costs are those associated with the characteristics of a particular day such as the difficulty of transportation to school or the excitement of a pep rally. Individual costs are those that are specific to the individual student like ability.

Shared Costs

Shared costs depend on the characteristics of the school and school day itself. For example, days adjacent to breaks or to weekends are more costly for students because students receive value from beginning the break early or they are tired after the break's end. Costs may also be higher later in the school year or semester due to the upcoming summer vacation or the stress of finals and term papers and less costly at the beginning of the year because students are excited to see friends and schoolwork has not yet become too difficult or demanding. Days at traditional schools with their stricter requirements and rigid schedules may also be more costly than days at alternative schools.

Let each day of school at a traditional school have a base shared cost $\bar{C}_t$ for all students, and let that cost be a function of all characteristics of the school day. Shared cost has a known distribution such that $\bar{C}_t \sim (\bar{C}, \sigma)$ where $\bar{C}$ is the average shared cost of a school day and σ is the standard deviation. $\bar{C}_t$ can be further broken down into components:

$$\bar{C}_t = \mathbf{X_t}\beta + E_t, \tag{2.1}$$

where X_t is a vector of day characteristics and E_t is an i.i.d daily cost shock.

Individual Costs

Individual costs are the costs that students face on a given day above or below the base cost of schooling for that day. Let the individual cost component be c_{it}. The individual cost can also be broken into components:

$$c_{it} = -b_i - hc_{it} + e_{it}, \tag{2.2}$$

where b_i is a measure of a student's ability that does not vary over the school year. b_i encompasses a student's aptitude for school, cognitive and non-cognitive skills that were accumulated before the year began, and any other factors that affect a student's time invariant cost of schooling. Individual costs are decreasing as ability increases. hc_{it} is a measure of accumulated human capital at time t such that:

$$hc_{it} = f(a_{i1}...a_{it-1}), a_{ij} = \{0, 1\} \tag{2.3}$$

a_{ij} is an indicator for whether student i attended school on day j. Human capital on date t is a function of students' attendance investment in all days of the school year before date t. Human capital increases in days attended and decreases in days truant: $\frac{\partial HC_t}{\partial a_{ij}} > 0$ for all $j < t$ and the effect of a past absence or truancy is allowed to decay overtime such that $\frac{\partial HC_t}{\partial a_{it-1}} \geq \frac{\partial hc_t}{\partial a_{it-2}}$. In other words, more recent truancies may have a larger impact on human capital than truancies further in the past.

Total Cost

The total cost faced by a student on date t is the combination of the shared and individual costs. Students choose each day whether or not to attend school based on the revealed cost of the day. The full cost becomes:

$$C_{it} = \mathbf{X_t}\beta - b_i - hc_{it} + \widetilde{e_{it}} \tag{2.4}$$

The cost a student faces on any given day is determined both by the characteristics of the day as well as the student's history of attendance and personal ability. For simplicity and ease of manipulation, all elements of the cost enter the model additively. However, this assumption will be relaxed in the empirical results to allow human capital to interact with other elements of the individual's daily cost, specifically, the level of human capital a student has will affect the opportunity cost change at 16. Students who have accumulated less human capital are likely to face a larger increase in the opportunity cost of schooling once they turn 16.

2.4.3 Returns to Education

The value that a student receives upon completing high school is based on the credential that she receives and her inputs into that credential. Students with lower human capital have lower returns from graduating with a traditional diploma and are more likely to receive an alternative diploma or drop out.

Returns are a function of accumulated human capital over the course of high school, individual ability, and credential quality.

$$R_{iT} = R(a_{i1}...a_{iT}, b_i, DEGREE) \tag{2.5}$$

where a_{it} is the truancy choice for individual i on date t and b_i is the time-invariant individual component defined the same way as above and T is the last day of high school. $DEGREE$ is the type of degree a student received if any. Students receive returns from schooling even if they do not complete a degree because those students still have accumulated human capital that has value in the labor market.[16] Returns are increasing in human capital, ability, and degree type quality.

2.4.4 Decision Problem

The Series of Decisions

Students make two types of decisions: the decision to begin a year of schooling and the decision to attend school each day during that year. When students decide to start a year, they do so with an expectation about the costs and benefits of the year and their ability to complete that year of schooling and subsequent years in order to get a degree. Students decide whether or not to begin 9th grade and upon entering 9th grade decide, each day, whether to attend school or be truant. In the summer after 9th grade, before the beginning of the next school year, students decide whether to return to their neighborhood school, transfer to another school, or dropout of school completely. If students choose to stay in school, upon entering 10th grade, students decide each day whether to attend

[16]Dropouts benefit in the job market from developing more skills while in high school even through they do not receive a credential (Tyler, 2004)

school or to be truant. Students repeat this process for grades 11 and 12 by deciding whether or not to begin the grade and then whether or not to attend or be truant on each day of those grades. At the end of 12th grade, students still in traditional school have either completed the requirements to graduate on time and receive a diploma or they have not.[17]

Decision Problem

Students believe the daily cost of schooling they will face is: $\widetilde{c_{it}} = \bar{C}_t - b_i + e_{it}$, in other words that accumulated human capital during the year will not change future costs, and they make decisions according to that expected cost ($\widetilde{c_{it}}$). At the beginning of each year while the student is in her original traditional school she chooses an attendance rule, attend if $c_{it} < c_i^*$, to satisfy the following condition:

$$\max_{a_{i1}...a_{iT}} E\left[R(a_{iy1}...a_{iT}, b_{iy}, DEGREE)\right] - E\left[\sum_{t=y1}^{T} \widetilde{c_{it}} a_{it}\right] \tag{2.6}$$

where a_{it} is the student's decision to attend on a specific date t, $y1$ is the first school day of school year y and T is the last day of high school. b_{iy} is the student's ability at the start of year y. The student has an expectation of how costs are distributed across school days at the beginning of the year, but the actual cost of each individual day will only be revealed to the student on that day. The day could have a particularly high cost due to its attributes, $\mathbf{X_{t-1}}$, such as being a Friday before a break, due to an idiosyncratic shock to the cost of the school day, E_{t-1}, such as a weather incident, or a personal shock to the student's daily cost, e_{it-1}, such as a broken car or family emergency.

[17]While it is possible for 12th grade students who have failed to return for the following year in their traditional school it is unlikely. Also, students who return would not be "on-time" graduates and are sometimes required to attend "adult education" rather than traditional school so the student may no longer have a choice over school type.

The attribute of this model that causes students to potentially miss more school than they ex-ante plan is that they do not account for the effect that truancy has on future costs. While the actual individual cost on a day is $c_{it} = -b_i - hc_{it} + e_{it}$, students believe the cost to be $\widetilde{c_{it}} = -b_i + e_{it}$. In other words, they believe that the average expected individual cost will be the same even if they miss a day of school. This belief on the student's part makes sense if she conceives of each missed day as an independent decision. Consider a student who has missed a day of school $(t-1)$ because of a high cost draw: $c_{it-1} > c_i^*$. This choice affects the student's cost distribution in the future. Beginning at date t, the student will face a slightly higher cost each day than she ex-ante expected because $hc_{it} < hc_{it-1}$ and $hc_{it} < E[hc_{it}]$ due to the missed day. This can result in the student missing additional days she did not plan to miss because her decision rule remains the same but $P(c_{it} > c_{it}^*)$ has increased. The high cost draw c_{it-1} that resulted in the student missing day $t-1$ could be due to any element of the cost function.

This increase in cost can cause truancy to snowball. After missing one day, the probability of missing any one day after that is higher, which will cause the student to miss more days she would have ex-ante expected which will in turn cause her to miss even more days after those days are missed. This result is consistent with the survey findings that a majority of dropouts report missing school so frequently before dropping out that they could not catch up. Figure 1 depicts a stylized example of this effect. Each day the low $c*$[18] student misses results in a greater increase in $P(c_{it} > c_{it}^*)$ as the cost distribution shifts to the right with each absence. The lower panel shows how truancy would snowball.

It is unlikely, to say the least, that high school students are mathematically

[18]A student with a lower cost threshold for missing school.

evaluating every decision they make about attending school. However, the logic that students follow a simple decision rule is intuitive. Students have a threshold for what they are willing to put up with at school and that threshold is affected by their ability, traits, and what they can expect to get out of high school. A simple decision rule for teenagers is also consistent with neuroscience research on adolescent brains. Adolescents are hypersensitive to immediate rewards and have more problems with inter-temporal decision making than adults (Blakewell & Robbins, 2012). The idea that a student wakes up each morning and decides if it is worth it to get on the bus based on what she expects of her day without considering the complicated long term consequences is consistent with this research. Students do learn in this model and can change their attendance plans, however they only change those plans between school years. This reflects the idea that students may make a "fresh start" at the beginning of the year and would take into account any damage they did to their ability to graduate and receive a credential the year before.

If students had complete knowledge of the education production function and were aware of the impact of human capital on the cost distribution, each student would attend as much as she planned to attend ex-ante. The decision rule c^* would be chosen to take into account a daily cost increase after a student misses a day of school. This could have two effects: first, students would set their decision rule with a higher c^* to accommodate cost increases in the future because they believe that the benefits to completing the year of schooling still outweigh this higher tolerance for daily costs. Second, it could result in students deciding that the higher cost threshold they would need to complete the year is too high to tolerate and choose not to enroll at all or to enroll in an alternative program.

This model demonstrates a process by which students who drop out get "off track" or "disengage" and the importance of viewing truancy and dropout as a dynamic process. A single high cost day or series of high cost days could start a chain reaction leading students to miss more and more school eventually dropping out. Students with an already low cost threshold for attending skip school because the revealed daily cost is too high and then begin to have chronic truancies that make it harder for them to return to school. These could even be students who ex-ante truly planned to finish school and had the ability to do so.

In addition to unplanned drop out, students could also have an unplanned transfer into a lower quality degree program because the shared cost of schooling, $\bar{C}_t$, depends on the school that the student attends. The average cost and returns of attending an alternative school are lower than attending a traditional school because alternative schools waive many of the requirements of traditional schools and have more flexible classes and schedules. A student who misses more days than she ex-ante planned in year y, and therefore has accumulated a lower level of human capital than she had planned, may decide not to continue in traditional school in year $y+1$ if the costs are too high. Consider a student who at the beginning of year y plans to to spend years y and $y+1$ in traditional school. In year y she misses more school than planned and, due to her lower than planned human capital, there is no attendance rule that would make the expected costs of year $y+1$ equal to the benefits such that:

$$E\left[R(a_{iy+1,1}...a_{iT}, b_{iy+1}, TRAD)\right] \geq E\left[\sum_{t=y+1,1}^{T} c_{it}^{\widetilde{TRAD}} a_{it}\right] \tag{2.7}$$

where $TRAD$ is the indicator for a traditional high school diploma program and $c_{it}^{\widetilde{TRAD}}$ is the individual's daily cost of attending a traditional high school. Instead of completely dropping out, the student may transfer to an alternative school in year $y+1$. Even though the returns to attending an alternative school are lower than a traditional school, if the costs are also low enough the student will transfer. The student will begin year $y+1$ at an alternative program if there exists an attendance rule, attend if $c_{it}^{ALT} \geq c_{it}^{*ALT}$, for the alternative school that satisfies:

$$E\left[R(a_{iy+1,1}...a_{iT}, b_{iy+1}, ALT)\right] \geq E\left[\sum_{t=y+1,1}^{T} \widetilde{c_{it}^{ALT}} a_{it}\right] \tag{2.8}$$

where ALT is the indicator for an alternative degree and $\widetilde{c_{it}^{ALT}}$ is the student's daily cost of attending an alternative high school. If there is no attendance rule for which the expected costs of the alternative school equal the expected returns, the student will drop out of school completely.

2.5 Identification and Empirical Strategy

2.5.1 Identification

In order to test the predictions of the model and evaluate the dynamic impact of costs on attendance and attendance on graduation. I exploit plausibly exogenous variation in the timing of a large opportunity cost increase that occurs when a student turns 16. Opportunity cost increases at 16 for students because they face more employment possibilities outside of school and less coercion to be in school. Although the state mandatory schooling age is 18, the district court will not hear truancy cases for students aged 16 and above nor will the district refer truancy cases to their own internal remediation system after 16. There is also a dramatic change in child labor laws. Before age 16 students can only work a maximum of 18 hours per week and only between the hours of 7am and 7pm during the school year. At 16 students are permitted to work 4 hours each school day and 8 hours on non-school days for a maximum of 48 hours per week (a 30 hour jump). Students aged 16 and older may also work as early as 5am and as late as 10pm on schooldays as well as hold additional types of jobs compared to younger students. Younger students are limited in job choice to low impact service jobs such as working at soda fountains. Because truancy rules and child labor laws are age based, students are subject to this increase in opportunity cost at different times of the year and at different points in their high school careers depending only on their birthdays. The majority of students turn 16 during their sophomore years of high school: 16% first semester, 40% second semester, and 20% during the summer before junior year. The remaining students turn 16 either during freshman year (3%), in the summer before sophomore year (4%) or during the first semester of junior year (17%).

The effect of work on high school students is well studied and there are mixed conclusions on its effect. Many studies show that students who work in high school have higher labor market returns later in life compared to similar non-working peers (Meyer & Wise, 1982; Ruhm, 1997; Light, 2001; Carr et al, 1996). However, work during the school year, especially intensive work of 15-20 hours per week has been found to have a negative effect on academic outcomes by lowering GPA (Greenberger & Steinberg, 1986; Eckstein & Wolpin; Oettinger, 1999) and decreasing the probability of completing high school (Stern, 1995). Studies that use child labor laws as an instrument for work find negative academic effects of work (Rothstein, 2007; Lee & Orazem, 2009; Schulenberg & Bachman). This is potentially because students aged 16 and over have much greater opportunities to do the type of intensive work that has been shown to have negative effects on academic performance.

The model predicts that an increase in opportunity cost will increase the probability that a student will be truant from school immediately following the cost increase at 16 and that probability will continue to increase with each subsequent day the student misses because of an increased probability of a cost draw above c^*. Consider two students, A and B, with identical ability, and before date t, identical attendance histories. Both students have the same attendance rule, c^*_{AB}. On date t student A turns 16 but student B does not turn 16 until 6 months later. Immediately student A faces a higher cost to each day of schooling. There is, therefore, a higher probability that student A will miss school on any given day compared to student B. Let the students' initial cost distribution be $c_{ABt} \sim (\bar{C_{AB}}, \sigma)$. After his birthday student A faces the distribution $c_{ABt} \sim (\bar{C_{AB}} + \gamma, \sigma)$ where γ is the additional opportunity cost of increased work opportunities and decreased truancy enforcement.

There should not necessarily be an immediate observable response in A's actions. Whether A's observable behavior changes relative to B's is determined by their c^*_{AB}, how long A has been subject to the additional cost and therefore how likely it is that he has received a sufficiently high daily cost draw to miss school, and how large that increase in average cost actually is. If A and B have a sufficiently high c^*_{AB} the increase in opportunity cost for A would not increase the number or frequency of his absences compared to those of B. This is borne out in the data as nearly 30% of students in the sample never have a truancy at any age. Figure 1 demonstrates the differential effects of an opportunity cost increase and truancy on high and low c^* students in the model. High c^* students are unlikely to experience cost draws high enough to result in truancy, and even if they do, the resulting increase in the probability of an additional truancy is very low compared to the effect on low c^* students. However, if students A and B have a low c^* they will be sensitive to the effect of the opportunity cost increase. Figure 2 illustrates the model's prediction of how the opportunity cost change at sixteen will affect a marginal student and the accelerated increase in truancy that can result.

This prediction of the model, that an increase in opportunity cost will not necessarily have an immediate effect on student behavior but will potentially result in a longer term truancy increase as they are exposed to the cost increase for longer, is integral to the empirical strategy. A standard regression discontinuity design would be unable to capture this long term effect. Comparing students just before and after turning 16 would result in an underestimation of the effect of the opportunity cost change despite beginning a process of increased truancy and potential dropout.

The main identification assumption, that students face a permanent increase in costs at age 16 the timing of which they cannot control, is tied to the assumptions in standard regression discontinuity and instrumental variable designs. The interaction of exact birth date and school entry laws has been used before to identify the effect of years of schooling on outcomes later in life (Dobkin & Ferreira, 2010; Smith, 2009; Elder 2010; McCrary & Royer, 2011; Evans et al, 2010) . However, the focus has been on the effect of an entire year age difference by comparing students who were just too young to enter school to those who were just too old. In that case, the students who were just too young and did not enter would be subject to mandatory schooling laws for a full year less than the just old enough students. Previous studies have found evidence that students who barely miss school entrance age drop out earlier (Angrist & Krueger, 1991; Bell et al, 2015). Students who are held back later in middle school and are thus older in high school are also more likely to dropout (Jacob & Lefgren, 2009). By examining how the age change affects students' truancy decisions during the year, I build on this literature by providing insight as to why this age timing matters by examining truancy and disengagement as the mechanism through which relatively older students drop out rather than the importance of binding mandatory schooling laws. I also expand the methodology to daily rather than yearly choices.

2.5.2 Daily Level Truancy Empirical Strategy

The main goal of the empirical strategy is to measure how a change in opportunity cost affects student's dynamic attendance decisions including, how the cost change immediately affects her truancy, how it affect her decisions as she spends more time subject to the higher cost, and how those decisions are related to dropout. I estimate a daily truancy equation in which truancy depends on the student's accumulated human capital, the student's time invariant ability, the average daily cost of school, and the student's age relative to 16:

$$\begin{aligned} Truant_{igt} =& \alpha_0 + \beta_0 * \mathbf{HC_{igt-1}} \\ & + \beta_1 * I(age > 16) + \beta_2 * (age_m - 16*12) + \beta_3 (age_m - 16*12)^2 \\ & + \beta_4 * (age_m - 16*12) * I(age > 16) + \beta_5 * (age_m - 16*12)^2 * I(age > 16) \\ & + \gamma_{ig} + \delta_t + e_{igt} \end{aligned} \tag{2.9}$$

$\mathbf{HC_{igt-1}}$ is the level of accumulated human capital on date $t-1$ in grade g for student i, γ_{ig} is the student's time invariant ability, δ_t is the date fixed effect for date t, and $\beta_1 - \beta_5$ are coefficints on a quadratic function of age_m which is age in months, centered around age 16, allowing for both a different slope and discrete change at age 16. $I(age > 16)$ is an indicator that the student's age in years is greater than 16 on date t and e_{it} is a student by day error term.

The empirical model is consistent with the motivating theory model. Student's daily attendance is determined by the opportunity cost they face from being above 16 or not, how long they have been subject to that cost increase, their accumulated human capital up until date t, their individual ability, and the shared daily cost.

The human capital function is flexible but my preferred base specification is a function of total truancies during the school year before date t, total truancies in the semester before date t and total truancies in the month before date t. This specification is nested, and the coefficients on the semester and month truancies are the additional effect that those more recent truancies have on the student's actions at date t beyond the effect of the total number of truancies that year.

$$\begin{aligned}\mathbf{HC_{igt-1}} = \beta_{0sch} * \sum_{j=schyear_0}^{t-1} Truant_{igj} + \beta_{0sem} * \sum_{k=semester_0}^{t-1} Truant_{igk} \\ + \beta_{0mth} * \sum_{l=month_0}^{t-1} Truant_{igl}\end{aligned} \tag{2.10}$$

Where $schyear_0, semester_0$, and $month_0$ are the first date in the school year, semester, and month that contain date t respectively, and $Truant_{igj}$ is the student's previous truancies in that year (j), semester (k), or month (l).

2.5.3 Between Year School Transitions

OLS

At the end of each school year students decide whether they will continue in their current school, transfer to another school, or drop out of school entirely. Students decide if the expected value of attending that year is higher than the expected cost. The student's truancies in the year determine the human capital she will go forward with the next year. This human capital accumulation determines the costs and benefits the student faces from continuing schooling. The baseline specification for this continuation decision is:

$$CONTINUE_{icg} = \beta_{0s} + \beta_1 * \mathbf{DEMOG_{icg}} + \beta_2 * \mathbf{SCORE_{icg}} \tag{2.11}$$
$$+\beta_3 * \mathbf{HC_{icg}} + \gamma_{cg} + e_{icg}$$

$CONTINUE_{icg}$ is an indicator for whether student i in cohort c in grade g continues in her present traditional school for grade $g+1$. $\mathbf{DEMOG_{icg}}$ is a vector of demographic characteristics for student i, $\mathbf{SCORE_{icg}}$ is a vector of student achievement scores: GPA, ELA score, and whether or not the student has to repeat math. γ_{cg} is a student by cohort (entrance year by entrance school group) fixed effect and e_{icg} is an error term. $\mathbf{HC_{icg}}$ is a truancy human capital specification.

The baseline human capital function is the total number of truancies the student has accumulated over the course of the year.

$$\mathbf{HC_{icg}} = \sum_{j=schyear_0}^{T_g} Truant_{icgj} \tag{2.12}$$

Where $schyear_0$ is the first day of the school year and T_g is the last day of the school year. $Truant_{icgj}$ takes on the value 1 if student i in cohort c and grade (school year) g was absent on date j. In this specification it is the total amount of time that students spent in school that matters. Because there is only one instrument, I use total truancies during the year as the measure of human capital in both the OLS and IV models. However, truancies are likely to have different effects depending on when they occur during the year and section 6.2.3 includes additional OLS estimates which provide suggestive evidence on effect of truancy timing on continuance.

IV approach

The OLS specification could suffer from omitted variables bias or endogeneity problems. Students who have obligations at home, dislike school, or have limited parental supervision are likely to both be absent more during the year and to not return the following year. In order to disentangle the causal effect of truancy on dropout, I use the change in opportunity cost at 16 as an instrument for the total number of truancies that a student has during the year. Students who have been subject to the higher opportunity cost longer will be induced to have more truancies than those who spend less time with the higher cost.

First Stage:

$$\begin{aligned}\mathbf{HC_{icg}} =&\alpha_1 * (AGE_{ig} - 16 * 12) + \alpha_2 * (AGE_{ig} - 16 * 12)^2 \\ &+ \alpha_3 * (AGE_{ig} - 16 * 12) * I(AGE_{ig} > 16) \\ &+ \alpha_4 * (AGE_{ig} - 16 * 12)^2 * I(AGE_{ig} > 16) + \alpha_5 OVER16_{ig} \\ &+ \alpha_6 * \mathbf{DEMOG_{icg}} + \gamma_{cg} + \eta_{icg}\end{aligned}$$

$$(2.13)$$

Second Stage:

$$CONTINUE_{icg} = \beta_{0s} + \beta_1 * \mathbf{DEMOG_{icg}} + \beta_2 * \widehat{\mathbf{HC_{icg}}} + \gamma_{cg} + e_{icg} \tag{2.14}$$

Here AGE_{ig} is the age of individual i at the end of school year g and $I(AGE_{ig} > 16)$ is an indicator that the student was over 16 at some point during that year. This specification is almost identical to the one used in the the daily level results. $OVER16_{ig}$ is an indicator for whether the student will be above 16 at the beginning of the next school year. $\mathbf{DEMOG_{icg}}$ is a vector of demographic characteristics including race, gender, school entrance age, and English as a second language and $\widehat{\mathbf{HC_{icg}}}$ is the predicted human capital value from the first stage. η_{icg} and e_{icg} are error terms. The identification assumption is that, when controlling for the age the student began high school, the time a student spends over 16 during the year only affects the probability of continuation the following year through the accumulation of human capital and whether the student will begin the following year after turning 16. For the instrument to not be valid, the amount of time a student is 16 would have to affect her ability and decision to continue school without affecting her truancy. One possibility is that students who are over 16 are more likely to have excused absences they would be more likely to not continue school because they missed important lessons even though they did not miss them due to truancy. However, that does not appear to be the case. Students can not excuse their own absences at 16 nor is there are measured change in the rate of excused absences once a student turns 16.

2.6 Results

2.6.1 Within Year Truancy

Baseline Results

The vast majority of students turn 16 either during their sophomore years or in the summer between sophomore and junior year of high school. Students who turn 16 during their freshman year were either held back at some point, or originally began school in a state with different entrance laws. Students who turn 16 during junior year are similarly young for their grade and likely began kindergarten outside of the state's public school system. I focus on the effect of turning 16 on truancy in 10th grade because that is when students who are of the correct age for their grade will turn 16. However, I also show results for 9th and 11th grade attendance as well and all students are 16 before the beginning of 12th grade.

Figure 3 provides a visible depiction of the effect of turning 16 on truancy. The top panel shows the effect on raw truancy using a fractional polynomial fit and the bottom shows the effect on residual truancy. Using raw truancies is problematic because age is correlated with time during the school year. If students are more likely to be truant at the end of the school year and more likely to be 16 at the end of the school, evidence that turning 16 causes truancy could be erroneous. In the second panel of Figure 3 average residual truancies are plotted against age. Residual truancy is the estimated residual from regressing truancy on date and individual fixed effects. The graphical depiction of residual truancy mimics the double fixed effect set up of the econometric results.

$$Truant_{igt} = \alpha_0 + \gamma_{ig} + \delta_t +_{igt} \tag{2.15}$$

Where γ_{ig} is the fixed effect for student i in grade g and δ_t is the fixed effect for date t. The residual truancy is $\hat{r_{igt}} = Truant_{igt} - \hat{Truant}_{igt}$ or the level of individual truancy not explained by the date and individual fixed effects.

The increase in truancy seen in Figure 3 is consistent with the empirical results. Table 3 contains the baseline estimation results for the effect of the opportunity cost change at 16 for students in grades 9, 10, and 11. Coefficients and standard errors have been scaled by 1000 for ease of reading. The bottom panel includes back of the envelope calculations of the relative effect of turning 16 on truancy after 1, 3, and 6 months in 10th grade. These percent increases are scaled by the average level of truancy: 2%. The main coefficients of interest are those on the interaction terms months 16*over 16 and months 16 squared*over 16 If both these coefficients are positive, then the longer a student is 16, the more likely she is to be truant, and that the frequency of truancy increases over time. For students who turn 16 during 10th grade this is the case. A student who has been 16 for 3 months in 10th grade is 1% more likely to be truant each day compared to other students. However, while this number seems very low, the relationship between age and truancy is quadratic; the longer a student has been subject to the higher opportunity cost after turning 16, the more likely she is to be truant. After 6 months a student is 4.4% more likely to be truant and after 12 months, which is the equivalent of barely missing the school entrance cutoff and being nearly a full year older than her youngest peers, is 14% more likely to be truant.

These results are also consistent with the predictions of the model. The model predicts that there should not necessarily be an immediate discrete change in student behavior upon turning 16 but that, once the opportunity cost increases, students who have a low enough cost rule will miss more and more days as they push their future costs higher.

There are slightly different patterns for students who turn 16 in 9th and 11th grade. Students who turn 16 in 9th grade are older and likely repeated a grade in the past. These students face even greater increases in the probability of truancy at 16. A student who has been 16 for 3 months in 9th grade is nearly three times as likely to be truant compared to a student who has not turned 16. Students who turn 16 in 11th grade are those who began school young. The coefficient on over 16 is positive and significant implying that students who turn 16 in 11th grade after most of their peers do appear to immediately change their behavior unlike students who turn 16 earlier. However, the point estimates on the interaction terms, though insignificant, are negative suggesting that, at most, the average student who turns 16 in 11th grade does not have worsening truancy over time.

The estimated effect of human capital, accumulated past truancies, on truancy is also consistent with the model. Students who have had truancies in the past are more likely to be truant especially if the student had a truancy in the past month. The human capital specification is nested so the effect of one truancy in the past month is the sum of the coefficients on year truancies, semester truancies, and month truancies. A student who has had only one truancy in 10th grade but in the last month is 40% more likely to be truant compared to the baseline. However, if that one truancy occurred the semester before and the student has had no subsequent truancies she is 0.5% less likely to be truant. The coefficient on month truancies. This is consistent with the model. More recent truancies would have a larger effect on the cost of the school day because the student has not had time to catch up to missed work or may be in the habit of not attending school. The effect of past truancies on human capital and future truancies does decay over time.

The students who have the largest increases in truancy after turning 16 are the students who have had past truancies. In terms of the model, these are students who have accumulated less human capital than their peers by the time they turn 16 and are also students who are likely to have a lower c^* so they miss lower cost days than other students. Additionally, students who have already become chronic truants are more likely to know when truancy enforcement changes so the opportunity cost increase for these students is higher at age 16.

The specification in table 4 explores the interaction between opportunity cost and human capital accumulation. It is essentially a triple difference specification which interacts the accumulated human capital a student has at her 16th birthday with turning 16. $TRUANT3$ is an indicator that the student has already had 3 truancies by the time she turns 16. These students would have already received letters from the school about the truancy remediation process and are more likely to know that enforcement changes at 16. The coefficients of interest are the interaction terms with $TRUANT3$, again, all coefficients have been multiplied by 1000 for ease of reading. The coefficients on all interaction terms are significant. For 10th graders, the coefficient on the interaction term with months 16 squared is negative, but very small indicating that, for chronic truants, there is a diminishing rate of increase in truancy frequency after 16. The bottom panel of the table includes similar calculations to the baseline results of the percentage increase in truancy, relative to the 2% average rate, for students who have fewer that 3 truancies and students with more than 3 when they turn 16. Although the coefficients are imprecisely estimated, students who have had fewer than 3 truancies before 16 are less likely to truant once they turn 16. However, students with previous truancies are much more likely to be truant. They are 34% more likely to be truant after 1 month, 63% after 3 months, and twice as likely to be truant after 6 months. The small average results in the baseline appear to be driven by chronic truants.

"At Risk" Students

Much of the dropout prevention literature focuses on students that are at high risk for dropout. These are generally students with low socioeconomic status, who go to low performing schools, and they are disproportionately likely to be racial minorities. Figure 4 plots the residual truancy for students at high risk (low performing) schools in the district compared to students at other schools. Students at low performing high schools are more likely to be truant than other students, on average students are truant 7 – 10 days freshman year depending on the school and up to 90% of students will have at least one truancy compared to less than 50% at high performing high schools. As can be seen in figure 4, high risk schools have more truancy overall and the rate of truancy may accelerate after 16 faster than for lower risk schools. Table 5 formalizes these results, the results in column (1) are similar to a triple difference specification. Interaction terms allow students at high risk schools to have different truancy age slopes before and after 16 compared to low risk school students. While students in high risk schools are more likely to be truant and have a steeper slope of truancy accumulation to age, the increase in that slope after 16 is not statistically significantly different from the increase for all students. Column (2) is estimated only for high risk schools. In this specification, only students at high risk schools contribute to estimating the day fixed effect reflecting the assumption that shared cost is school specific. None of the results are economically or statistically significantly different from those for all students.

Race and gender are also often used to identify higher risk students and it is possible that Black and Hispanic students will have different truancy behavior compared to other students after turning 16. Figure 5 plots the residual truancy for Black and Hispanic students compared to other students. While both Black and Hispanic students have higher rates of truancy, there does not appear to be much differential effect of turning 16.[19] Figure 6 also shows that there is little difference in truancy behavior at or before 16 for female and male students.

Robustness Checks

One important threat to identification is that teenagers may not actually face an increase in opportunity cost at age 16. Students could already be working off the books when 15 or could not work at all at any age. They could also be unaware of both the change in labor law and truancy policy. It is also possible that the apparent increase in truancy after age 16 is an artifact of the data. One advantage of the double fixed effects specification is that it controls for each date separately which rules out the possibility that age is a proxy for season or time during the school year. An additional benefit of this design is that it captures the

[19]Econometric results for Black and Hispanic students will be made available in an online appendix or upon request.

dynamic effects of opportunity cost over time. Using a regression discontinuity design to compare students just before and after 16, would falsely conclude that there is no difference in their behavior which would imply that either, students do not react to the cost change, or that there is no cost change.

There is also the possibility that there is nothing special about turning 16, and that students' behavior changes in a similar fashion at all of their birthdays. Figure 7 compares the changes in truancy at ages 15.5 and 16. There is no visibly discernible change between the fractional polynomial estimated from 14.75 to 15.5 and from 15.5 to 16. In fact, when attempting to include indicators in the baseline regression for over 16 and over 15.5 simultaneously the two were co-linear. Students who turn 15.5 in 10th grade and who turn 16 are extremely similar. They have both transitioned from 9th to 10th grade, and they are both within the normal school starting age.

Table 6 compares three different estimates of the increase in truancy at 16 taking into account that there could also be a change in behavior at age 15. Columns (1) and (2) are the baseline 9th and 10th grade estimates, and column (3) is a specification that includes both indicators for time over 16 and time over 15 simultaneously for 9th graders. The coefficients on the over 16 variables interaction variables are robust to this inclusion with the exception of the over 16 indicator. The baseline estimate for the linear and squared interaction terms are -9.51 and 6.83 respectively for 9th graders and with the inclusion of the over 15 interactions those become -10.5 and 6.7. These estimates are within the 95th percentile confidence interval of each other.

Another possibility is that truancy isn't actually increasing at 16 and the apparent increase is just absences changing labels. Students may have more autonomy when they turn 16 and, as a result, absences that parents would have excused when they were younger now become truant absences. In other words, students are in school the same amount of time, but being out of school is called something different. This is inconsistent with the data. Figure 8 plots residual absences and tardies, defined the same way as residual truancy, and shows virtually no change in the probability of excused absence at age 16. There seems to be no evidence that students are substituting excused absences with truancies once they turn 16. Additional suggestive evidence is found in the bottom panel of Figure 8, there is an increase in residual tardies at age 16. Tardies, like truancies are within control of the student rather than the parents. Table 7 compares the estimated effect of turning 16 on truancy, absence, and tardies. The coefficients of interest are the interaction terms of $months16 * over16$ and $months16squared * over16$ Each column is a separate regression and the outcomes are truant, absent, and tardy. Unlike the coefficients in the truancy regression which are positive, the coefficients on the interaction terms for both tardies and excused absences are negative.

2.6.2 Between Year School Transitions

OLS Baseline Results

There are four possible outcomes for a student transitioning to 10th, 11th, or 12th grade. The student can continue in her current traditional school, drop out of school entirely, transfer to an alternative school, or transfer out of the district. For the following results I focus on the decision of students to stay in their traditional high school but will also discuss the results for continuing in any school or dropping out. As can be seen in Figure 9, students with higher levels of truancy in 10th grade are less likely to continue in their neighborhood school between 10th and 11th grade and less likely to continue in any school. A student with three truancies compared to none is nearly 10 percentage points less likely to continue in traditional school. While most students who have more than 30 truancies (15% of the school year) do not return to their neighborhood school the following year, there are still students with very high truancy levels that do return as can be seen in the right tail of the graph. Even some of the students that miss more than an entire semester worth of classes return for the following year. From comparing the top and bottom panels of the figure, it is clear that most of the students who leave their traditional school do not immediately drop out but rather transfer to another school. Even the majority of students with a semester's worth of absences continue into 11th grade.[20]

Table 8 provides the OLS results for the effect of truancies on the transition from 10th to 11th grade. Each additional truancy decreases the probability of returning to school by .006 to .01. Overall 82.5% of tenth graders continue in their original school so each truancy decreases the probability of returning by 0.5 to 1 percentage point. Controlling for demographics and cohort, each additional truancy makes a student 1% less likely to continue in their neighborhood school. Columns 3 -5 include additional achievement covariates including the students' test scores, whether they have to repeat the same math class the next year, and two different measures of GPA: weighted GPA and an indicator for having a GPA under 2.0 (a 2.0 GPA is a requirement for graduation). The direction of the truancy effect is the same with the addition of these covariates but the magnitude decreases. Including the test score and GPA variables is prob-

[20]Patterns are similar for the 9th to 10th and 10th to 11th grade transitions and will be made available in an online appendix and upon request.

lematic because both test scores and GPA are highly correlated with truancy. Truancy could cause low grades when students miss lessons, and low grades could cause truancy by making students dislike school. Table 9 expands the set of covariates to include excused absences as well. The coefficients on truancies in all specifications are robust to the inclusion of excused absences. The coefficient on truancy is also larger than the coefficient on excused absences, each truancy is associated with a larger decrease in continuance than an excused absence. The results for the transition from 9th to 10th and 11th to 12th grade follow a similar pattern.[21]

While my focus is on the effect of truancy on continuing in traditional high school, there is also an important effect on continuing in any high school and dropping out. Tables 10 and 11 include the estimates for the effect of truancy on continuing in any school and dropping out respectively. Each cell is a regression and the estimates are the coefficient on truancy. Truancy decreases the probability of attending any school by between 0.1 and and 0.6 percentage points which is much smaller than the effect on continuing in traditional school. The effect on dropout is also much smaller in part because most students transfer to another program before dropping out.

IV

While the OLS results are informative, the change in the magnitude of the truancy variable when covariates are added demonstrates the potential for omitted variable bias and endogeneity concerns in the OLS model. The total effect of truancy on transition could be higher if a student's truancy caused her bad

[21] Results will be made available in an online appendix and upon request

grades or caused the necessity of repeating math. Table 12 contains results for the IV specification for the transitions between 9th and 10th and 10th and 11th grade. The effect of truancy on continuation in traditional school is estimated to be much larger in this specification compared to the OLS estimates. With each additional truancy a student is 2 percentage points less likely to continue in her neighborhood public school. With an average continuation rate of 83%, a chronic truant with only 3 truancies, is 10% less likely to continue in her neighborhood school compared to a student with no truancies.

In panels 2 and 3 I include the results for the effect of truancies on continuing in any school and dropping out. Like the OLS results, these estimates are much lower than the estimates for continuing in traditional school Truancy has a larger effect on continuing in traditional school than on dropping out entirely. This is likely because most students first transfer to an alternative school and then drop out. The difference between the OLS and IV estimates highlights the complicated role truancy plays in dropout. Truancy not only affects continuation directly, but also through its effect on grades and test scores. This interaction is difficult to tease out in OLS.

OLS: Alternative Human Capital Specifications

In the IV specification above the only human capital measure used is the sum of all truancies during the year because there is only one instrument: the amount of the year the student was 16. However, it is likely that not all truancies should affect school continuation in the same way. Truancies could be more or less important later in the year or there could be a non-linear relationship between truancy and continuation. Table 13 allows truancy to have a non linear relationship with school continuance, specifically that any truancy beyond the 3 that label a student as a chronic truant could have an aggravating or dampening effect on continuance. The results show, across specifications and grade transitions, that each truancy beyond the first three decreases continuance by less than each of the three initial truancies. This diminishing effect of truancy is likely due to the district's policy focusing on 3 truancies as an important benchmark for students.

Finally, truancies could have different effects on continuation depending on the time of year in which they occur. Table 14 repeats the baseline OLS analysis allowing the coefficient on truancy to vary month to month. As is predicted by the model and documented in the literature, truancies that occur in the late spring, May and June, have the largest effect on continuing school. This could be due in part to the fact that some of those truancies are students who have already dropped out but are still counted as truant. Despite the importance of May and June, the effect of monthly truancies changes by grade. In 10th grade truancies in September and December have the largest effect on continuing of any month other than June, a pattern not seen in 9th grade. In 11th grade, truancies in September have a larger effect than truancies in any other month but June. The importance of these fall semester truancies could provide early warning indicators for dropout.

2.7 Conclusion and Policy Implications

The important implication of these results is that the changes in costs that students face in attending school can have dramatic effects over time when left un-addressed. While an increase in the daily cost of schooling may have either no immediate effect, or a very small immediate effect, there is potential for the effect to increase overtime and cause students to make more and more bad decisions endangering their ability to graduate. This is especially true for students who already have histories of poor attendance. The more truancies a student has had in the past, the more likely she is to be truant again and the truancies that are most predictive are those within the same month. Additionally, students who have already been labeled chronic truants (they have 3 or more truancies) by the time they turn 16, accumulate truancies at a faster rate than their peers

These results support the idea that the most effective programs for preventing dropout, once students are in high school, are those that lower the cost of physically attending school by increasing truancy enforcement or decreasing the negative aspects of school. Many of the high school dropout prevention programs that have the best evidence for success are high intensity mentorship programs that serve to both lower the psychic costs of school for a student by providing a social relationship and by helping students keep up with course work and monitoring her attendance (Bell et. al., 2015; Oreopoulos et al, 2014; DeSocio et al, 2007).

The results also cast doubt on the idea that it is “too late” to intervene with older students. Students react to a change in the cost of schooling as late as their junior year of high school and that cost increase over time decreases their likelihood of remaining in school. Some of the students who eventually drop out have even passed their high school exit exams before leaving school. These results are consistent with evidence that older students do benefit from enforced attendance. In college mandatory attendance policies aimed at students not much older than high schoolers have been shown to be effective at increasing learning when they increase attendance (Marburger, 2006).

The results are also additional evidence that aggressive truancy prevention strategies, such as case worker intervention, are more effective than less intense strategies like warning letters (Lemieux et al, 2011). Students are more likely to be truant after enforcement has decreased once they turn 16 and can no longer be referred to the county's juvenile court or the district's own intensive remediation system. This is especially problematic for students that already have chronic truancy problems because they are likely to notice the drop in enforcement. Sustaining efforts to keep students in school when they are mandated to be there, no matter how old the student is or how late in her career she begins missing school, could help prevent this disengagement and eventual dropout.

The results also call for further study to understand the role alternative high school programs play in the dynamic process of dropout. More than half of students who transfer to an alternative school in the sample eventually drop out and the transfer students have lower human capital in their original school before transferring compared to those who stay. While these schools award lower quality diplomas, it is unclear from the data whether the net effect of alternative schools is positive or negative. Students would be worse off in alternative schools if they would have been able to finish in a traditional high school if an alternative school were not available or if they were not pushed toward one. However, students would do better in an alternative school if the student would have dropped out earlier if alternative school were not an option. One important challenge of alternative programs, especially those with large online learning components, is to prevent he structure and content of alternative programs from re-enforcing bad behavior on the part of students. When students with truancy problems are transferred to online programs with minimal oversight, their truancy and disengagement could easily get worse.

2.8 Tables

Table 2.1: Student Sample by Year of Schooling

Initial School	Entrants n	Enter Dropout	Exiters n	Exit Dropout	Traditional Graduate	Traditional Dropout	Alternative Graduate	Alternative Dropout	Transfer Out
1	1,800	.088	1,390	.045	.642	.017	.089	.087	.165
2	1,650	.139	1,400	.110	.478	.060	.038	.098	.327
3	3,200	.036	2,660	.014	.706	.009	.080	.048	.155
4	2,570	.155	2,160	.119	.497	.070	.049	.102	.281
5	2,300	.085	1,790	.043	.577	.013	.078	.106	.226
6	2,000	.033	1,740	.011	.764	.007	.089	.037	.104
7	1,470	.111	1,560	.082	.507	.036	.066	.084	.308
8	1,540	.099	1,260	.050	.614	.025	.084	.090	.187
9	3,200	.040	2,880	.020	.754	.005	.042	.051	.148
10	2,100	.106	1,640	.074	.622	.038	.061	.094	.185
11	2,500	.080	1,860	.054	.559	.032	.056	.121	.233
12	2,600	.066	2,090	.033	.681	.019	.089	.065	.145
13	3,170	.165	2,700	.132	.483	.068	.071	.104	.274
14	3,000	.027	2,670	.009	.785	.005	.055	.032	.121
15	2,730	.046	2,270	.017	.618	.011	.066	.049	.256
16	2,300	.034	2,030	.012	.754	.008	.067	.036	.134

Entrants is the number of students who begin high school in that school and exiters is the number of students who also finish high school in that school. "Enter Dropout" is the fraction of students who enter the initial school and eventually drop out and "Exit Dropout" is the fraction of students who drop out for whom their entrance school was their only school. "Traditional" graduates and dropouts are students who begin and end their high school careers in a traditional district public high school, "alternative" graduates and dropouts are students who transfer within the district to an alternative school program before either graduating or dropping out. "Transfer" students are students who transfer out of the district so their graduation status is unknown.

Table 2.2: District Demographics and Graduation Patterns

Demographic Variables	Sample Mean	Traditional Graduate	Traditional Dropout	Alternative Graduate	Alternative Dropout	Transfer Out	Freshman Truancies
Black	.117	.528	.024	.058	.078	.313	5.2
White	.252	.730	.009	.075	.032	.154	1.5
Hispanic	.440	.545	.042	.077	.111	.226	5.7
Asian	.172	.810	.012	.033	.035	.113	1.5
Female	.485	.614	.022	.069	.060	.191	3.7
Male	.515	.614	.030	.066	.085	.206	4.0
Black Female		.558	.022	.062	.072	.268	4.5
Black Male		.50	.025	.053	.084	.3383	4.3
White Female		.739	.007	.074	.024	.156	1.4
White Male		.722	.011	.076	.039	.157	1.7
Hisp. Female		.580	.035	.077	.092	.215	5.1
Hisp. Male		.510	.048	.076	.129	.237	5.6
Asian Female		.810	.008	.039	.025	.117	1.5
Asian Male		.780	.016	.031	.046	.124	1.8

Averages are calculated off the 9th grade initial sample. Racial categories are coded by the district, Hispanic, Black, White, and Asian are mutually exclusive categories. Traditional graduates are those that receive a diploma from their original school, traditional dropouts drop out without transferring first, alternative graduates transfer then graduate, alternative dropouts transfer then drop out, transfer out means the student leaves the district for another.

Table 2.3: Baseline: Daily level truancy for full sample (Coefficients scaled by 1000)

	(9th) Truant	(10th) Truant	(11th) Truant
over 16	3.44	-.218	2.27
	(1.62)	(0.277)	(0.760)
months 16	-11120	9873	383
	(36.18)	(77.68)	(65.1)
months 16*over 16	-9.51	0.0537	-0.895
	(2.62)	(0.120)	(1.51)
months 16 squared	0.0493	0.0103	0.379
	(0.00200)	(0.00518)	(0.628)
months 16 squared*over 16	6.83	0.0213	-0.332
	(0.877)	(0.0107)	(0.628)
year truancies	0.0331	-0.0734	-0.00110
	(0.0143)	(0.0159)	(0.0174)
semester truancies	0.876	0.808	1.28
	(0.0126)	(0.0166)	(0.0221))
month truancies	7.11	8.82	12.0
	(0.0548)	(0.0693)	(0.0869)
1 Month Effect	3.8%	-0.6%	–
3 Month Effect	182%	0.9%	–
6 Month Effect	–	4.3%	–
Observations	6810498	5588743	4648138

Standard errors are in parenthesis and are clustered at the cohort (entrance year by school) level. Each column is an individual grade, all students are 16 before 12th grade so 12th grade is omitted. Months 16 is the student's age in months centered around 16 and over 16 is an indicator that the student is over 16 on date t. Year, semester, and month truancies are the total number of truancies a student has accumulated before date t during that year, semester, or month. All specifications include both student level and date level fixed effects. Coefficients and standard errors have been multiplied by 1000 for ease of reading. The 1 month, 3 month, and 6 month effect is the percentage increase in truancy, relative to the baseline average of 2% for students after they have been 16 for 1, 3, or 6 months.

Table 2.4: Daily Truancy Interacted with Human Capital (Coefficients scaled by 1000)

	9th truant	10th truant	11th truant
over 16	-0.904	-0.718	0.340
	(.86)	(0.292)	(0.770)
months 16	-11110	9805	383
	(36180)	(77670)	(65.1)
months 16*over 16	-4.04	-0.402	-1.07
	(3.03)	(0.129)	(1.51)
months 16 squared	0.0493	0.0101	0.404
	(0.00200)	(0.00518)	(0.628)
months 16 squared*over 16	2.86	0.0282	-0.364
	(1.02)	(0.0116)	(0.628)
over 16*TRUANT3	17.1	4.44	18.5
	(3.74)	(0.694)	(1.28)
months 16*over 16*TRUANT3	-20.0	3.70	1.92
	(6.04)	(0.288)	(0.149)
months 16 squared*over 16*TRUANT3	14.8	-0.127	0.0245
	(2.01)	(0.0237)	(0.00659)
Year Truancies	0.0324	-0.0741	-0.00189
	(0.0143)	(0.0159)	(0.0174)
Semester Truancies	0.874	0.788	1.20
	(0.0126)	(0.0166)	(0.0221)
Month Truancies	7.11	8.81	11.9
	(0.0548)	(0.0693)	(0.0870)
Not Truant 1 Month Effect	–	-5.4%	–
Not Truant 3 Month Effect	–	-8.6%	–
Not Truant 6 Month Effect	–	-10%	–
Truant 1 Month Effect	–	34%	–
Truant 3 Month Effect	–	63%	–
Truant 6 Month Effect	–	101%	–
Observations	6810498	5588743	4648138

Standard errors are in parenthesis. Months 16 is the student's age in months and over 16 is an indicator that the student is over 16 on date t. Year, semester, and month truancies are the total number of truancies before date t during that year, semester, or month. "TRUANT3" is an indicator variable that the student had 3 accumulated truancies at the time she turned 16. All specifications include both student level and date level fixed effects. Coefficients and standard errors have been multiplied by 1000 for ease of reading. The 1 month, 3 month, and 6 month effect is the percentage increase in truancy, relative to the baseline average of 2% for students after they have been 16 for 1, 3, or 6 months. "Truant" indicates that the increase is calculated for students who had 3 or more truancies before turning 16.

Table 2.5: At Risk Schools (Coefficients scaled by 1000)

	(1) truant1	(2) truant1
o16	-0.00386	-1.45
	(0.306)	(1.00)
months 16	9842	38880
	(77480)	(281700)
months over 16	0.197	-0.115
	(0.134)	(0.423)
months 16 squared	0.00643	0.0324
	(0.00570)	(0.0187)
months over 16 squared	-0.00168	0.00469
	(0.0121)	(0.0354)
over 16 * in at risk school	-1.62	
	(0.724)	
months 16 * in at risk school	2.29	
	(0.179)	
months over 16 * in at risk school	-0.190	
	(0.308)	
months 16 squared * in at risk school	0.0264	
	(0.0134)	
months over 16 squared * in at risk school	0.0125	
	(0.0261)	
Year Truancies	-0.0736	-0.0947
	(0.0159)	(0.0337)
Semester Truancies	0.778	0.818
	(0.0166)	(0.0365)
Month Truancies	8.75	8.00
	(0.0693)	(0.155)
Observations	5588743	1000146

Standard errors in parentheses

Standard errors are in parenthesis and are clustered at the cohort (entrance year by school) level. Each column is an individual grade, all students are 16 before 12th grade so 12th grade is omitted. Months 16 is the student's age in months centered around 16 and over 16 is an indicator that the student is over 16 on date t. Year, semester, and month truancies are the total number of truancies a student has accumulated before date t during that year, semester, or month. All specifications include both student level and date level fixed effects. Coefficients and standard errors have been multiplied by 1000 for ease of reading.

Table 2.6: The Estimated Effect of Turning 15 (Coefficients scaled by 1000)

	(9th) Truant	(10th) Truant	(9th) Truant
over 16	3.44	-0.218	0.71
	(1.62)	(0.277)	(1.7)
months 16	-11120	.873	-11100
	(36180)	(77680)	(36170)
months 16*over 16	-9.51	0.0537	-10.5
	(2.62)	(0.120)	(2.60)
months 16 squared	0.0493	0.0103	0.02
	(0.00200)	(0.00518)	(0.002)
months 16 squared*over 16	6.83	0.0213	6.70
	(0.877)	(0.0107)	(0.912)
over 15			1.04
			(0.3)
months over 15			-0.2
			(0.12)
months over 15 squared			.091
			(0.012)
Observations	6810498	5588743	6810498

Standard errors are in parenthesis and are clustered at the cohort (entrance year by school) level. Each column is an individual grade, all students are 15 before 11th and 12th grade so they are omitted. Months 15 is the student's age in months centered around 15 and over 15 is an indicator that the student is over 15 on date t. Year, semester, and month truancies are the total number of truancies a student has accumulated before date t during that year, semester, or month. All specifications include both student level and date level fixed effects. Coefficients have been multiplied by 1000 for ease of reading.

Table 2.7: Baseline Model: Additional Outcomes (Coefficients scaled by 1000)

	(10th) Truant	(10th) Absent	(10th) Tardy
over 16	-0.218	1.02	0.98
	(0.277)	(0.425)	(0.53)
months 16	9873	-23870	31360
	(77680)	(119000)	(147300)
months 16*over 16	0.0537	-0.578	-0.018
	(0.120)	(0.185)	(0.23)
months 16 squared	0.0103	0.016	-0.007
	(0.00518)	(0.008)	(0.009)
months 16 squared*over 16	0.0213	-0.0137	-0.012
	(0.0107)	(0.0164)	(0.02)
Observations	5588743	5588743	5588743

Standard errors are in parenthesis and are clustered at the cohort level. Each column is an individual grade, all students are 16 before 12th grade so 12th grade is omitted. Months 16 is the student's age in months centered around 16 and over 16 is an indicator that the student is over 16 on date t. Year, semester, and month truancies are the total number of truancies a student has accumulated before date t during that year, semester, or month. All specifications include both student level and date level fixed effects. Coefficients have been multiplied by 1000 for ease of reading.

Table 2.8: Yearly Transition Results OLS: 10th Graders Continuing in Original School for 11th Grade

	(1) Continues	(2) Continues	(3) Continues	(4) Continues	(5) Continues
Total Truancies	-0.0103	-0.0102	-0.0110	-0.00607	-0.0067
	(0.0002)	(0.000206)	(0.00028)	(0.00215)	(0.00029)
ELA Score			0.0008		-.000027
			(0.00004)		(.000046)
GPA				0.0829	0.0785
				(0.0032)	(0.0037)
GPA under 2.0				-0.0740	-0.0720
				(0.0068)	(0.0067)
Demographics		X	X	X	X
Cohort FE	X	X	X	X	X
Observations	30542	30534	29531	30534	29531

Standard errors are clustered at the cohort (school by entrance year) level. Demographic variables include race, ethnicity, gender, ELL status, and high school entrance age. Students who transfer outside the district before the end of 10th grade (midyear) are excluded from the sample.

Table 2.9: Yearly Transition Results OLS: 10th Graders Continuing in Original School for 11th Grade with Excused Absences

	(1) Continues	(2) Continues	(3) Continues	(4) Continues	(5) Continues
Total Truancies	-0.0099	-0.0971	-0.0103	-0.00599	-0.00648
	(0.0002)	(0.000206)	(0.00028)	(0.00215)	(0.000290)
Total Absences	-0.00514	-0.00544	-0.00500	-0.00313	-0.00648
	(0.000256)	(0.000257)	(0.00026)	(0.000253)	(0.000290)
ELA Score			0.0008		-.000003
			(0.00004)		(.000046)
GPA				0.0766	0.0708
				(0.0033)	(0.0037)
GPA under 2.0				-0.0758	-0.0748
				(0.0068)	(0.0067)
Demographics		X	X	X	X
Cohort FE	X	X	X	X	X
Observations	30542	30534	29531	30534	29531

Standard errors are clustered at the cohort (school by entrance year) level. Demographic variables include race, ethnicity, gender, ELL status, and high school entrance age. Students who transfer outside the district before the end of 10th grade are excluded from the sample. Total absences are the total excused absences that a student has in addition to truancies.

Table 2.10: Yearly Transition Results OLS: All Grades Continuing in Any School

	(1) Any	(2) Any	(3) Any	(4) Any	(5) Any
9th to 10th	-0.00360	-0.00345	-0.00163	-0.00315	-0.00127
	(0.000078)	(0.000079)	(0.000092)	(0.000088)	(0.000101)
10th to 11th	-0.00446	-0.0043	-0.00286	-0.00396	-0.00247
	(.00008)	(.000083)	(0.0001)	(.00009)	(.00109)
11th to 12th	-0.00653	-0.0061	-0.00475	-0.00546	-0.00378
	(.00117)	(.00013)	(.00018)	(.000136)	(0.000194)
Demographics		X	X	X	X
Cohort FE	X	X	X	X	X
Test Scores			X		X
GPA measures				X	X

Standard errors are clustered at the cohort (school by entrance year) level. Demographic variables include race, ethnicity, gender, ELL status, and high school entrance age. Students who transfer outside the district before the end of 11th grade are excluded from the sample.

Table 2.11: Yearly Transition Results OLS: All Grades Dropout

	(1) Dropout	(2) Dropout	(3) Dropout	(4) Dropout	(5) Dropout
9th to 10th	0.00232	0.00204	0.000315	0.00229	0.00031
	(0.00031)	(0.0000317)	(0.0000329)	(0.000035)	(0.00035)
10th to 11th	0.00168	0.00135	0.000498	0.00146	0.00048
	(.000028)	(.000028)	(0.000033)	(.000031)	(.000035)
11th to 12th	0.00266	0.00185	0.000875	0.00187	0.000768
	(.0000412)	(.0000437)	(.0000572)	(.000046)	(0.000061)
Demographics		X	X	X	X
Cohort FE	X	X	X	X	X
Test Scores			X		X
GPA measures				X	X

Standard errors are clustered at the cohort (school by entrance year) level. Demographic variables include race, ethnicity, gender, ELL status, and high school entrance age. Students who transfer outside the district before the end of 11th grade are excluded from the sample.

Table 2.12: Yearly Transition Results IV: 9th to 10th and 10th to 11th Transitions

	Traditional		Any School		Dropout	
	9th to 10th	10th to 11th	9th to 10th	10th to 11th	9th to 10th	10th to 11th
Truancies	-0.0217	-0.0187	-0.00379	-0.00640	0.00307	0.00321
	(0.00313)	(0.00388)	(0.00091)	(0.0016)	(0.000732)	(0.00076)
female	-0.0103	-0.00563	-0.00237	-0.0036	0.0014	-0.00018
	(0.00384)	(0.00426)	(0.00136)	(0.00166)	(0.00067)	(0.00053)
Black	-0.0683	-0.0759	-0.0111	-0.0019	-0.0044	-0.00567
	(0.0112)	(0.0112)	(0.00348)	(0.0043)	(0.0002)	(0.00165)
Hispanic	0.00812	-0.0255	0.00471	0.00240	-0.0059	-0.0046
	(0.00742)	(0.0085)	(0.00265)	(0.0035)	(0.0015)	(0.0015)
ELL	0.0266	-0.0136	0.00471	0.00604	-0.0029	-0.0054
	(0.0102)	(0.0126)	(0.00313)	(0.00487)	(0.00217)	(0.002)
Observations	36797	29531	36797	29531	36797	29531
F-statistic	22.79	18.45	22.79	18.45	22.79	18.45

Instruments are a quadratic spline in age. Standard errors are in parenthesis and first stage F-statistics are reported.

Table 2.13: Yearly Transition Results OLS: All Grades Continues in Neighborhood School with Chronic Truancy

	(1) Continues	(2) Continues	(3) Continues	(4) Continues	(5) Continues
9th to 10th Total Truancies	-0.0217 (0.00217)	-0.0204 (0.00217)	-0.0193 (0.00214)	-0.0093 (0.00215)	-0.0113 (0.00214)
9th to 10th Truancies >3	0.0127 (0.00214)	0.0114 (0.00214)	0.0099 (0.00209)	0.00361 (0.00211)	0.005 (0.00208)
10th to 11th Total Truancies	-0.0304 (.00235)	-0.0261 (.00237)	-0.0239 (0.00233)	-0.0096 (.0023)	-0.0116 (.00229)
10th to 11th Truancies >3	0.0199 (0.00232)	0.0157 (0.00234)	0.0127 (0.0023)	0.00351 (0.0226)	0.00485 (0.0022)
11th to 12th Total Truancies	-0.0201 (.00213)	-0.0167 (.00212)	-0.0177 (.00207)	-0.00327 (.00203)	-0.00638 (0.00202)
11th to 12th Truancies >3	0.0096 (0.00211)	0.00625 (0.002)	0.00476 (0.002)	-0.00290 (0.00199)	-0.00113 (0.00194)
Demographics		X	X	X	X
Cohort FE	X	X	X	X	X
Test Scores			X		X
GPA measures				X	X

Standard errors are clustered at the cohort (school by entrance year) level. Demographic variables include race, ethnicity, gender, ELL status, and high school entrance age. Students who transfer outside the district before the end of 11th grade are excluded from the sample.

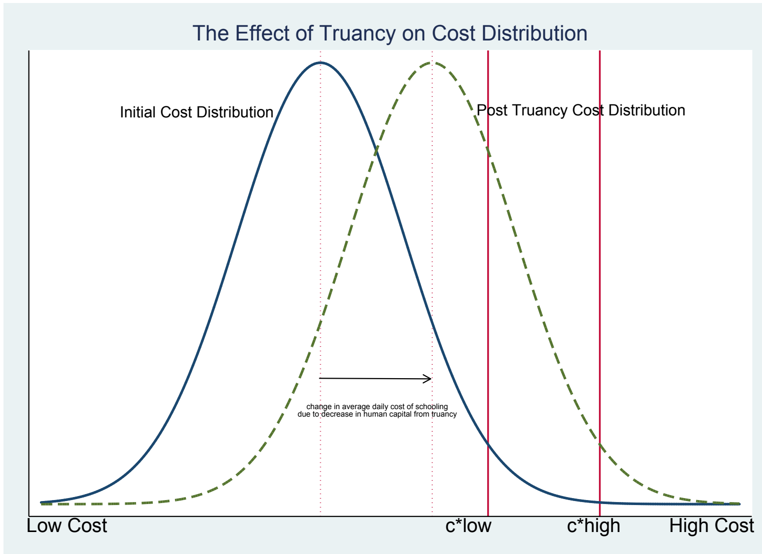
Table 2.14: Grade 10 to 11 – Estimates for Truancies by Month

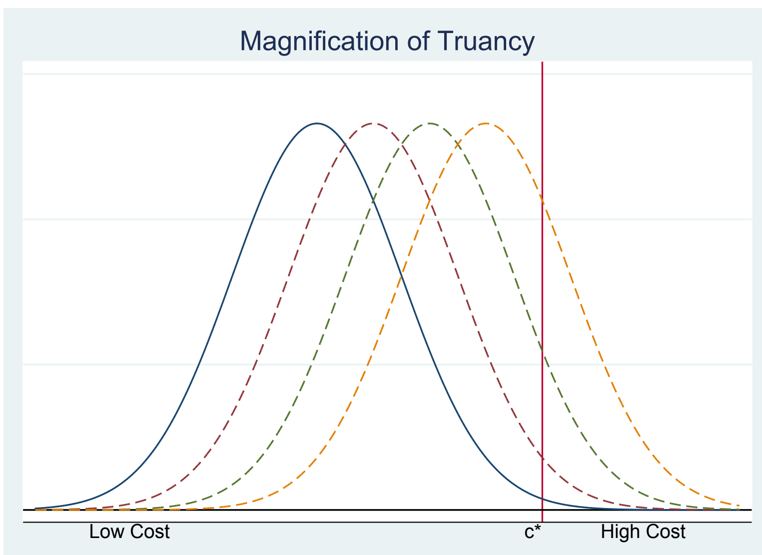
	(1) continues	(2) continues	(3) continues	(4) continues	(5) continues
September Truancies	-0.0176	-0.0152	-0.00630	-0.00908	-0.00207
	(0.00290)	(0.00291)	(0.00322)	(0.00280)	(0.00314)
October Truancies	-0.00539	-0.00601	-0.00158	-0.00278	0.000162
	(0.00237)	(0.00239)	(0.00269)	(0.00230)	(0.00262)
November Truancies	-0.00751	-0.00790	-0.00546	-0.00138	-0.000579
	(0.00283)	(0.00284)	(0.00310)	(0.00273)	(0.00302)
December Truancies	-0.0203	-0.0196	-0.0175	-0.0150	-0.0129
	(0.00274)	(0.00274)	(0.00291)	(0.00264)	(0.00283)
January Truancies	-0.00627	-0.00630	-0.00456	-0.00316	-0.00126
	(0.00207)	(0.00207)	(0.00223)	(0.00199)	(0.00218)
February Truancies	-0.00151	-0.00219	-0.00111	-0.000289	0.00205
	(0.00240)	(0.00239)	(0.00268)	(0.00230)	(0.00261)
March Truancies	-0.00952	-0.00948	-0.0144	-0.00647	-0.00966
	(0.00204)	(0.00204)	(0.00226)	(0.00196)	(0.00221)
April Truancies	-0.00657	-0.00629	-0.00880	-0.00391	-0.00549
	(0.00204)	(0.00204)	(0.00222)	(0.00196)	(0.00217)
May Truancies	-0.0124	-0.0121	-0.0199	-0.00895	-0.0150
	(0.00186)	(0.00185)	(0.00202)	(0.00178)	(0.00198)
June Truancies	-0.0236	-0.0228	-0.0251	-0.0117	-0.0169
	(0.00237)	(0.00236)	(0.00241)	(0.00228)	(0.00235)
Demographics		X	X	X	X
Cohort FE	X	X	X	X	X
Test Scores			X		X
GPA measures				X	X
Observations	30542	30534	29531	30534	29531
Adjusted R^2	0.096	0.102	0.108	0.171	0.152

Standard errors are clustered at the cohort (school by entrance year) level. Demographic variables include race, ethnicity, gender, ELL status, and high school entrance age. Students who transfer outside the district before the end of 10th grade (midyear) are excluded from the sample.

2.9 Figures

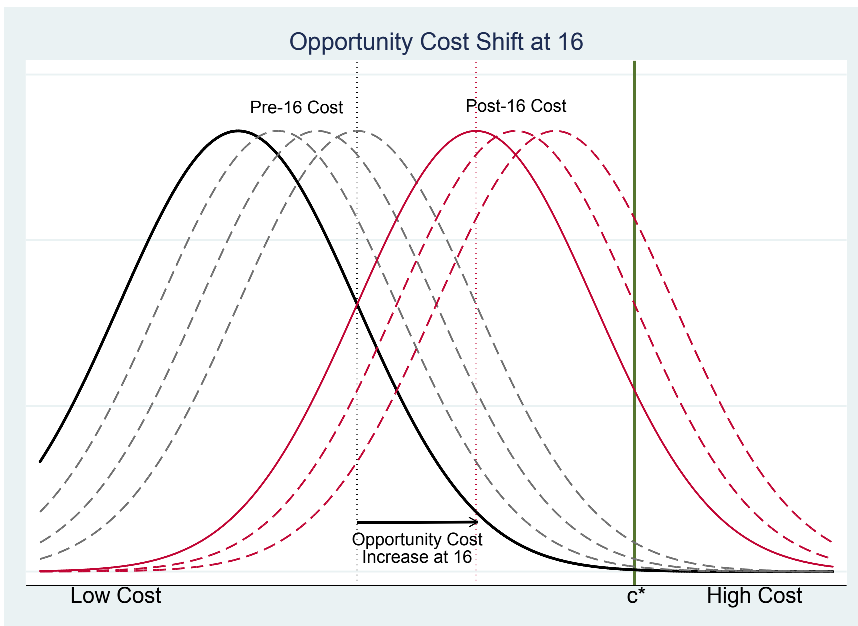
Figure 2.1: Theory Model Cost Distribution Change for High and Low C* Students





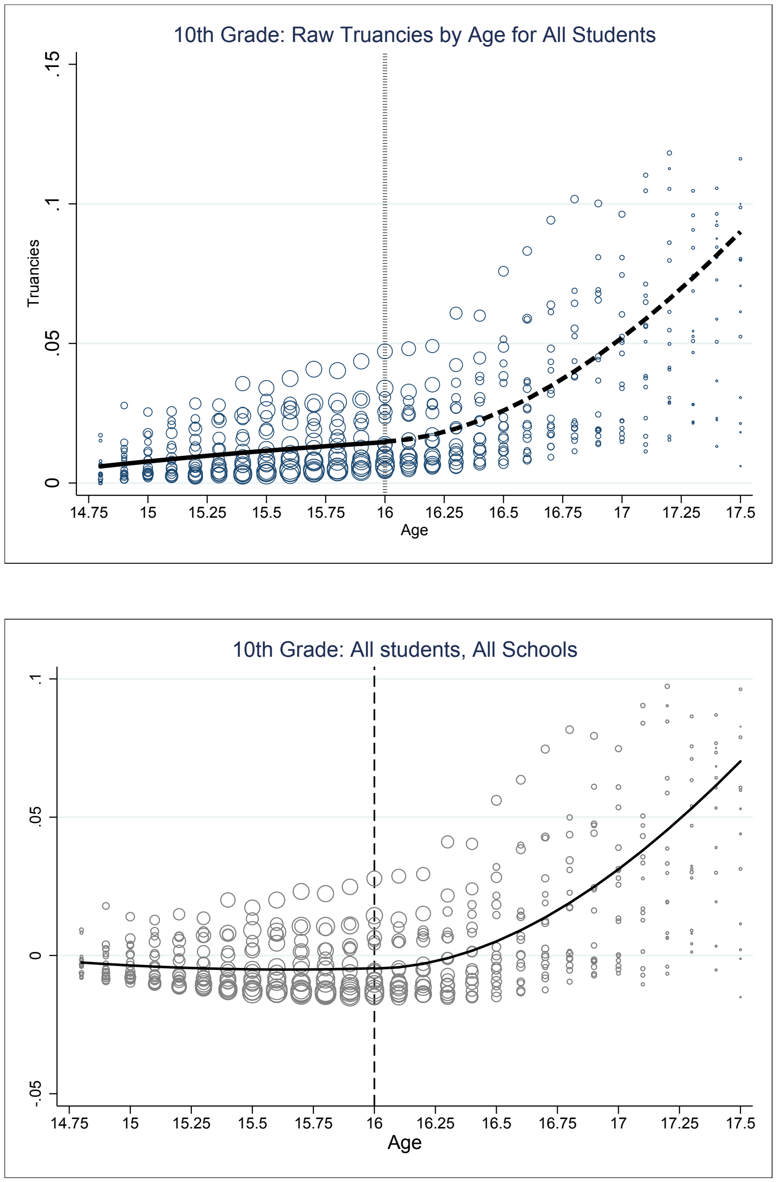
The top panel depicts a stylized version of the effect of truancy on future schooling costs and additional truancy. Students who receive a cost draw in the area below the initial cost distribution curve to the right of the c* line will select not to attend school. This decision lowers the student's human capital for subsequent days which increases the daily cost of schooling and shifts the cost curve from the initial distribution to the post truancy cost distribution. The probability of a subsequent truancy is now higher. Students with lower c* both have a higher likelihood of receiving a cost draw that causes them to miss school and face a higher increase to that probability for a given change in cost. The lower panel depicts how that effect magnifies.

Figure 2.2: Theory Model Cost Distribution: Effect of turning 16



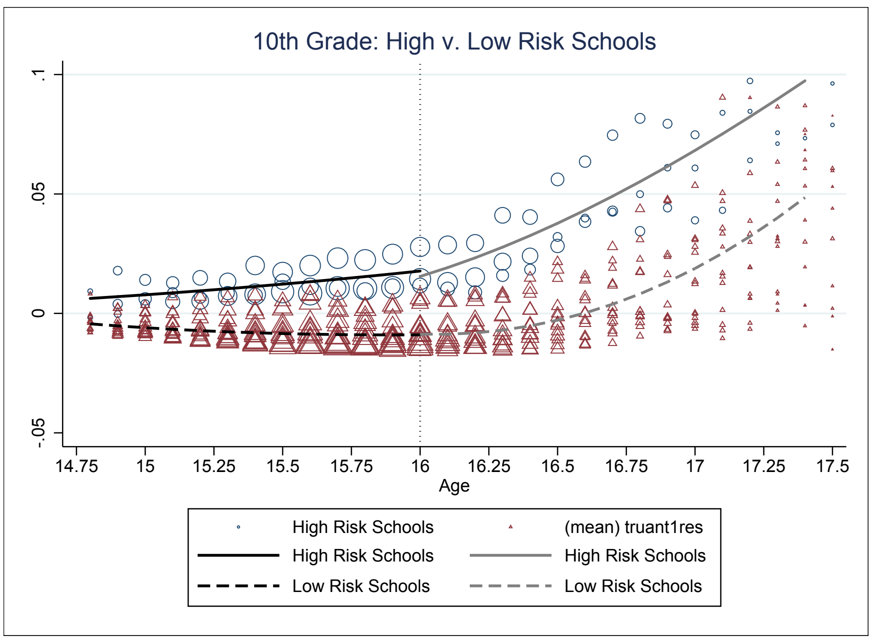
Students who receive a cost draw in the area to the right of the c* line from the initial (solid and black) cost distribution will elect to miss school. The probability of missing a subsequent day in now higher and the cost curve will shift right. With each subsequent truancy the cost curve shifts further right increasing the probability of another truancy at an increasing rate. After the student turns 16 there is an additional large increase in the daily cost of schooling which shifts the cost distribution to the right even if the student has no additional truancies. The permanent cost increase increases the rate of truancy accumulation for students who continue to be truant.

Figure 2.3: 10th Grade Truancy for All Students in All Schools



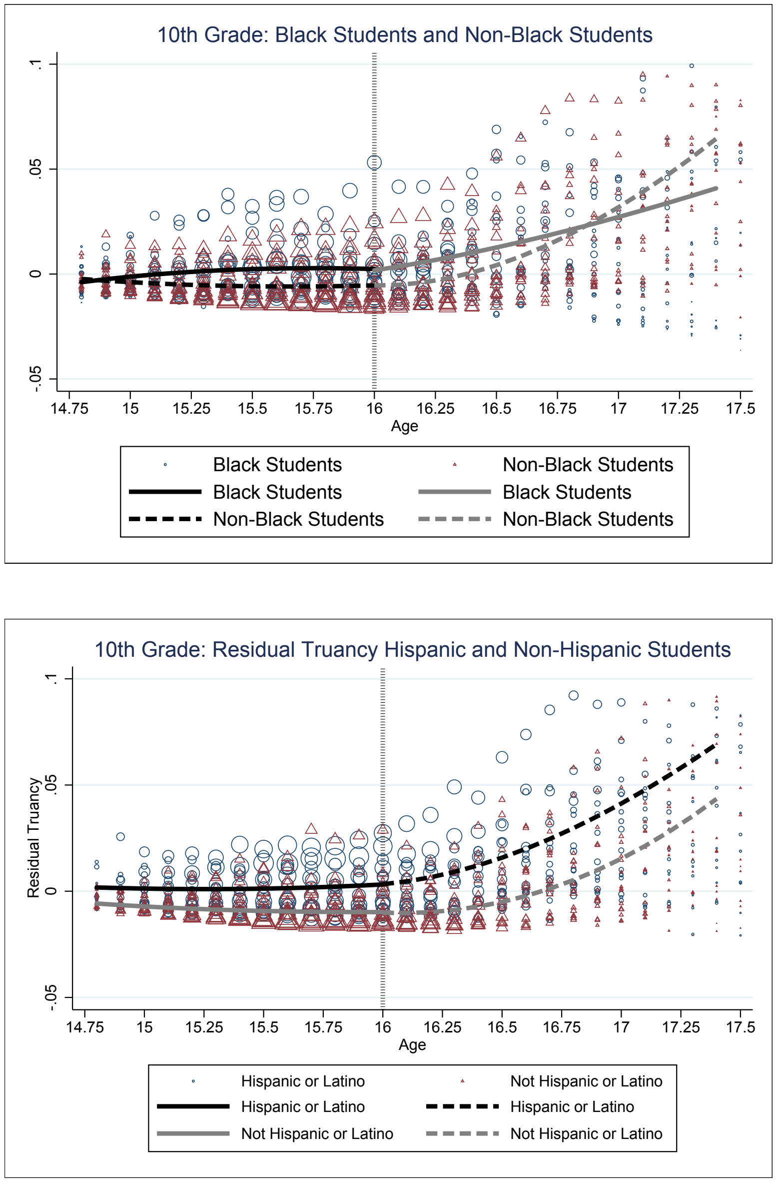
The top panel includes raw truancy by school and age cohort for all 10th grade students in the sample. Each circle is weighted by the number of students in that cell. The bottom panel includes residual truancy by the same school and age cohorts. Residual truancy is the estimated residual of an OLS regression of truancy on student and date fixed effects only.

Figure 2.4: Differences in Truancy Probability at 16 for Low and High Risk Schools



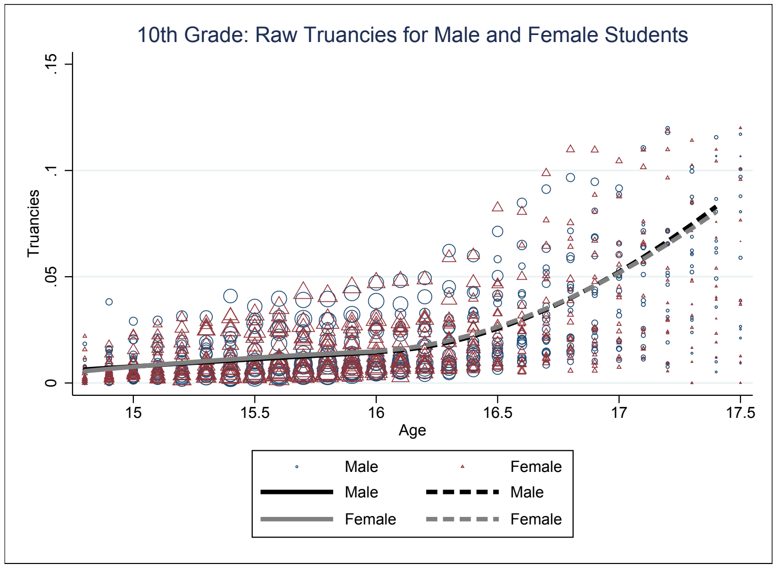
Each circle is the average residual truancy for an age group school cohort at a low performing high school weighted by the number of students in that school of that age. The triangles are the same grouping for higher performing schools. Residual truancy is the estimated residual of an OLS regression of truancy on student and date fixed effects only.

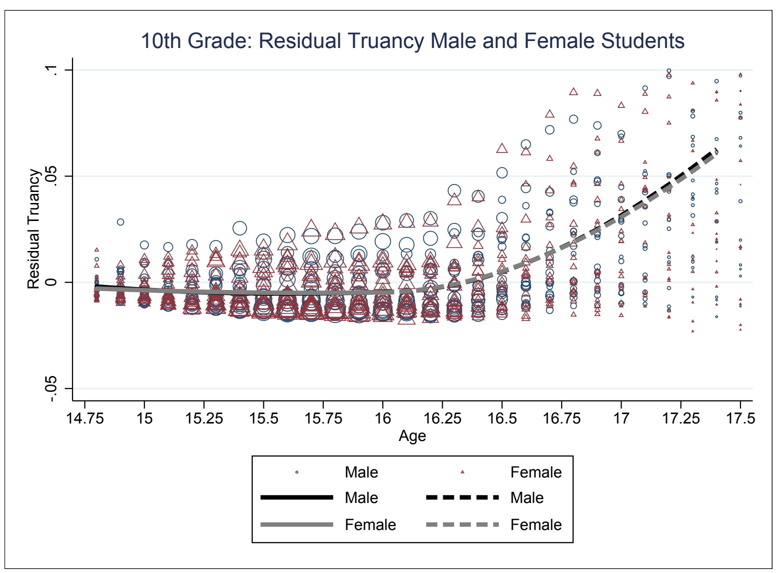
Figure 2.5: Black and Hispanic Students



Each circle is average residual truancy for an age group school cohort for Black students in the top panel and Hispanic students in the bottom panel. Each triangle is an age group school cohort for the non-Hispanic or non-Black students. markers are weighted by the number of students in that cell. Residual truancy is the estimated residual of an OLS regression of truancy on student and date fixed effects only.

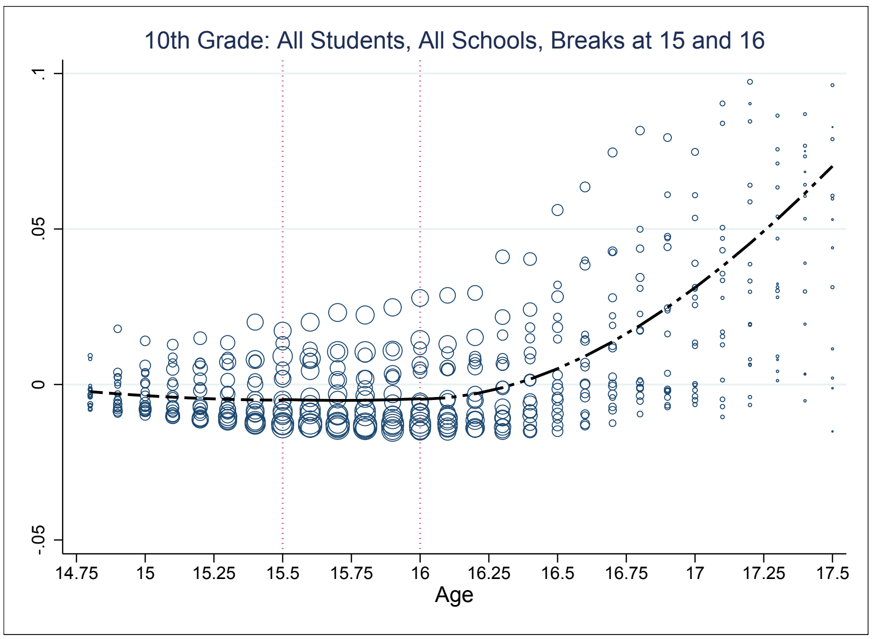
Figure 2.6: Male and Female Students





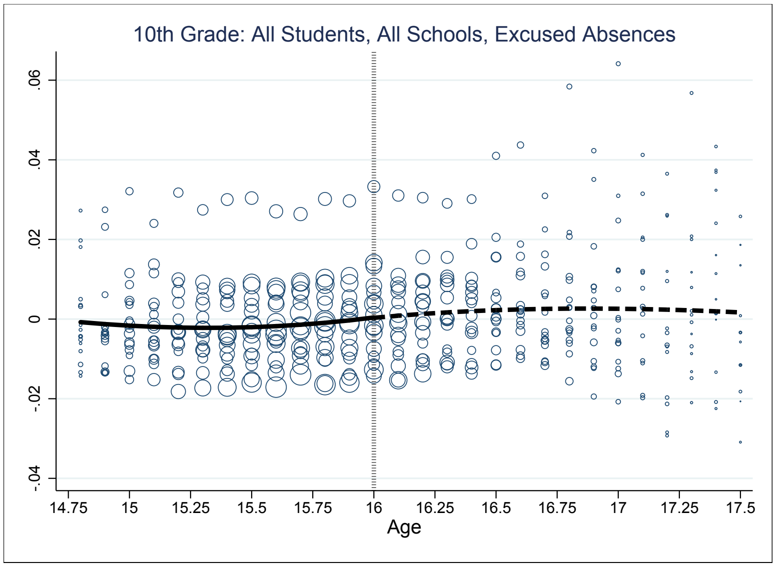
Each circle in the top panel is average raw truancy by age group school cohort for Male students, triangles are Female student averages. Markers are weighted by the number of students in that cell. The bottom panel is average residual truancy by the same groups. Residual truancy is the estimated residual of an OLS regression of truancy on student and date fixed effects only.

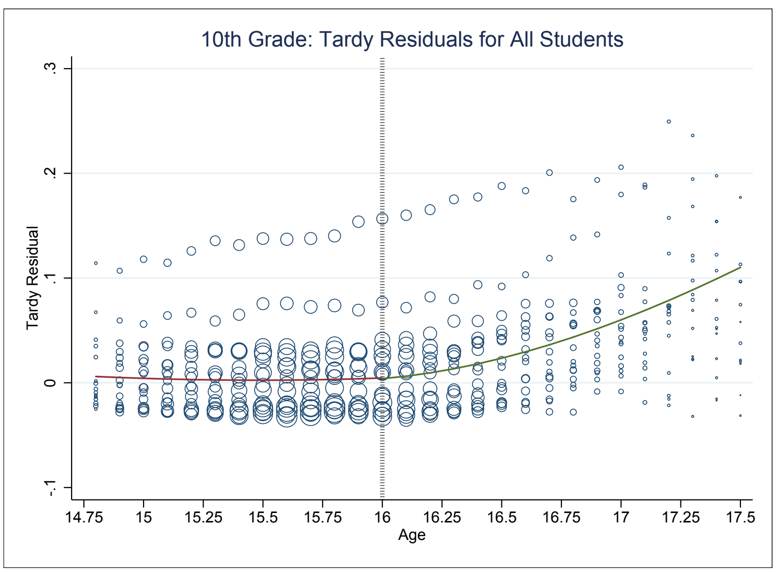
Figure 2.7: 15.5 vs. 16 Robustness Check



Each circle is the average residual truancy in an age group school cohort for all students. Markers are weighted by the number of students in that cell. Residual truancy is the estimated residual of an OLS regression of truancy on student and date fixed effects only. The flexible polynomial fit varies between 14.75 and 15.5, 15.5 and 16, and 16 and 17.5.

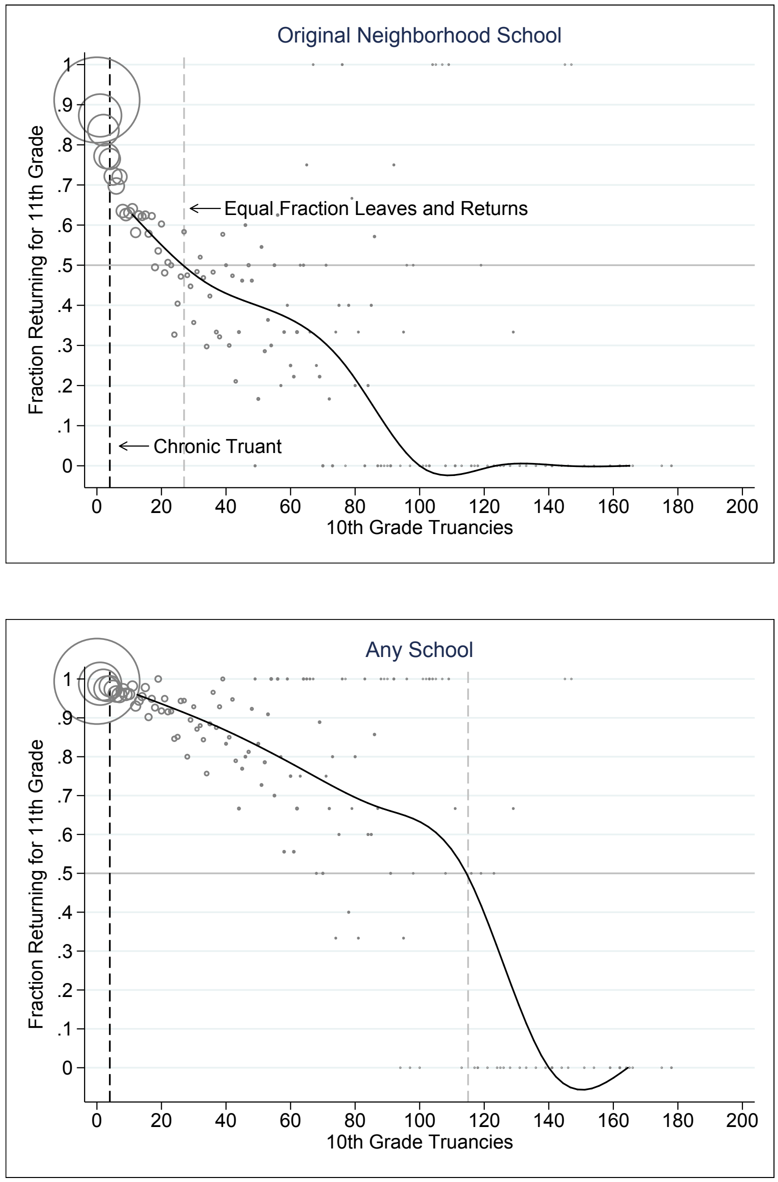
Figure 2.8: Tardies and Excused Absences





The top panel circles are average residual excused absences by school, age cohort weighted by cell size. The bottom panel circles are average residual tardies with the same grouping. Residual excused absences and tardies are the estimated residuals of an OLS regression of excused absences or tardies on student and date fixed effects only.

Figure 2.9: 10th to 11th Grade Transition



Each circle is weighted by the number of students in the sample at each school with that number of truancies, the line is a median spline also weighted by number of students with each level of truancy. Vertical lines indicate the level at which students are considered chronic truants and the median spline estimate of the number of truancies when a student has an equal probability of returning or leaving her school. The top panel is the return to the original school, the bottom panel is the return to any school.

CHAPTER 3

DOES ACCESS TO HEALTH CARE AFFECT TEEN BIRTH RATES AND SCHOOL DROPOUT RATES? EVIDENCE FROM SCHOOL-BASED HEALTH CENTERS (WITH MICHAEL LOVENHEIM AND RANDALL REBACK)

3.1 Introduction

Access to affordable health care for low-income Americans has become a preeminent policy issue in the U.S. The massive expansions of Medicaid and the State Children's Health Insurance Program that occurred over the past several decades have caused the gap in health insurance coverage between children from low-income and high-income families to all but disappear. Yet, health care access for children depends both on the affordability of care and on convenient availability of effective health care. Despite the elimination of the health insurance coverage gap, low-income families still face considerably higher costs of accessing high-quality health care services that drive disparities in the quality of care across the socioeconomic distribution (Smedley, Stith and Nelson 2003; Andrulis 1998). This quality gap can be attributed in part to supply-side factors, such as medical practices choosing not to accept Medicaid insurance and the reluctance of many doctors to locate their practices in low-income urban or rural areas. The gap also may be due to demand-side factors, such as low-income adults lacking information on appropriate health care providers or finding it difficult to take time away from hourly-paid jobs in order to accompany their children to these providers.

Medicaid eligibility leads to better health (Currie and Gruber, 1996a; Finkelstein et al., 2012; Kaestner, Joyce and Racine 2001; Currie, Decker and Lin 2008), more stable household finances (Gross and Notowidigdo, 2011) and higher educational attainment and earnings (Cohodes et al. forthcoming; Brown, Kowalski and Lurie 2014). However, inadequate access to primary care facilities and doctors among low-income families may preclude them from realizing these benefits of health insurance, which can render the roughly $86 billion the U.S. spends on Medicaid for children less effective. Given the large and persistent disparities across the socioeconomic distribution in academic achievement, health care access, and health status,[1] understanding how primary care health care services affect important life outcomes among youth is of high policy relevance.

In this paper, we explore whether expanding teenagers's access to health care influences their fertility rates and their educational attainment. We estimate the effects of providing primary care health services to teens through school-based health centers (SBHCs), which are health clinics located in a school or on school grounds. While they vary in size and scope, virtually all SBHCs provide basic preventative health services to students, and many of them also provide reproductive health services and contraception. SBHCs target underserved communities by predominantly locating in schools in low-income urban and rural areas. They therefore can reduce the costs of obtaining health care services for children from low-income families. Particularly for reproductive health among teenagers, these SBHCs may be extremely effective at increasing health care utilization because they reduce any reliance on parents to bring teenage students to the doctor. Currently, there are over 2,000 SBHCs in the US, and their prevalence has increased markedly over the past 25 years (see Figure 1). Although

[1] For example, see Currie, Decker and Lin (2008), Adler and Rehkopf (2008), Case, Lubotsky and Paxson (2002), Cunha et al. (2006), Conti, Heckman and Urzua (2010), and Todd and Wolpin (2007).

these centers are an increasingly important provider of primary care health services to youth in low-income areas, little is known about how they affect student health and education.

Our analysis makes two contributions to the literature. First, we present new evidence on the effect of primary health care services delivered through schools on teen birth rates. Whether a teenager gives birth is a critical health outcome that can have long-run consequences for the individual. Teen fertility rates in the U.S. are very high relative to similarly-industrialized nations but also have declined substantially in the last 25 years (Kearney and Levine 2012).[2] Currently, there is very little understanding of which policies are effective in reducing teen births. Providing health care services to teens, and in particular easy-to-access contraception through health centers in schools, may be an effective policy tool with which to lower teen birth rates. This paper is the first in the literature to estimate the causal effect of such primary care services on teen fertility.[3]

Second, our paper is the first to examine how primary health care services affect the educational attainment of children from low-income families. Providing access to primary health care services could increase educational attainment through any effect on child health as well as on family finances. A sizable amount of work has demonstrated that poor health or adverse health events among children are associated with worse long-run outcomes (e.g., Currie et al.

[2]Kearney and Levine (2014) find evidence that the MTV show *16 and Pregnant* explains about one third of the decline in teen births that occurred between 2009 and 2010. Due to the timing of when this show began, they are unable to examine the causes surrounding the large drop in teen fertility between 1990 and 2009, which is the period on which our study focuses.

[3]Much prior research has examined the effect of the diffusion of the birth control pill in the 1960s and 1970s on fertility decisions and life outcomes of somewhat older women (Goldin and Katz 2002; Bailey 2006, 2010; Ananat and Hungerman 2012). This literature does not analyze the effect of access to contraception among teens on fertility nor does it examine the efficacy of providing contraceptive services through schools, which is what we focus on in this analysis.

2010; Case, Fertig and Paxson 2005; Case, Lubotsky and Paxson 2002). Studies have found positive effects from specific types of child health interventions, such as hookworm eradication (Bleakley 2007), malaria eradication (Bleakley 2010) and school-based deworming drug interventions (Miguel and Kremer 2004). Several papers also have explored the 'fetal origins' hypothesis and have found evidence that pre-natal health care and health outcomes affect subsequent academic performance and success (e.g., Almond and Currie 2011; Figlio et al. 2014). Yet, we are unaware of prior research that credibly estimates the causal effect of comprehensive health services for school-age children on their educational attainment in an industrialized country setting.

A major hurdle in estimating the effect of health care services on fertility and education that has impeded prior research is that access to such services is not exogenously assigned: unobserved factors correlated with the quality of health care service availability are likely to be correlated with underlying fertility and education outcomes. We overcome this problem by exploiting the timing of expansions of school-based health centers in different school districts in the U.S. We obtained data from surveys of SBHCs conducted by the National Alliance on School-Based Health Care in 1998, 2001, 2004, 2007 and 2011. Centers are followed longitudinally, and in addition to being able to link them to the districts they serve, we have information on when each center opened, its size in terms of students served, hours open, staffing hours, and the specific health services it provides to students. We focus on centers that serve high school students (grades 9-12); overall, we observe 2,586 centers during our analysis period.

To identify the effect of SBHCs on teen fertility and dropout rates, we combine the NASBHC survey data with county-level information on births as well as district-level information on high school completion. For births, we use U.S. vital statistics data for which the smallest level of geographic identification is the county. Our main analysis focuses on births among 15-18 year old women, as they are most likely to have been recently enrolled in high school.[4] We measure treatment by whether there is *any* SBHC open in the county or school district as well as by treatment intensity using the primary care staff hours per week and total medical staff hours per week offered by *all* SBHCs in the county or district. These measures provide a comprehensive depiction of the medical services offered to students. As discussed further below, the process of opening an SBHC is typically initiated by hospital administrators who may then spend several years searching for a school to partner with, securing funding, and renovating a space to meet health clinic regulations. The timing of center entry varies significantly across counties and school districts as a result.

There are three potential threats to identification of the causal effects of SBHC services on teen fertility and dropout rates. First, the timing of center entry might be endogenous. In theory, this could bias estimates in either direction; centers might be opening when local officials are relatively resourceful, or when they are worried about unusually high rates of teen pregnancy or high school dropouts. Event study analyses provide extensive evidence that the timing of the initial center entering in a county or school district is not endogenous with respect to pre-treatment trends in our outcomes of interest. We therefore exploit the variation in timing across counties and school districts in initial center entry to identify how SBHC services affect teen outcomes. Second, yearly service level variation after initial entry might be endogenous. Event study analyses suggest this is indeed the case: services hours are targeted to areas that are experienc-

[4]We refer to birth rates among women aged 15-18 as "teen birth rates" throughout this analysis.

ing higher birth rates, especially right after an initial center opens. We address this issue by estimating an instrumental variables model that uses information about the first center opening in a district/county to predict future service level variation in that district/county. Third, there might be omitted variables, contemporaneous policies or shocks affecting outcomes in the low-income communities where SBHCs locate. All of our analyses control for state-by-year fixed effects, so state-level policy changes and state-level shocks are not a concern. Robustness checks add controls for various types of year-specific income categories; these robustness checks confirm that the main results are not influenced by omitted variables differentially affecting low-income populations. Falsification tests also indicate that coincidental policies or shocks are not a source of concern: we do not see fertility effects for women in their early 20s when health centers opened in local high schools, and we do not find any relationship between SBHCs and per pupil expenditures in schools.

Our findings suggest that SBHCs reduce teen fertility, with relatively large reductions among younger teens, African American teens, and Hispanic teens. Our baseline estimates show that center entry in a county reduces the teen birth rate by 1.3 per 1,000, which is a 3.0% reduction relative to the baseline birth rate. Just using the existence of a center in a county ignores potentially-important service differences across centers. Our preferred estimates examine the effects of changing the primary care or total medical staff hours offered by SBHCs; to facilitate interpretation, we scale the treatment effects to reflect the impact of adding services equivalent to an average-sized SBHC. These results indicate that service changes equivalent to opening an average-sized center lead to a 2.4%-2.7% reduction in births per 1,000 women aged 15-18. The IV results are larger: in our preferred model an average-sized center reduces teen birth rates by over 5%. We prefer the IV estimates because they address the endogeneity concerns related to SBHC service level variation as well as any attenuation bias from measurement error. Further analysis provides suggestive evidence concerning which types of services are most important for reducing teen fertility. The largest effects come from the subset of SBHCs that offer on-site prescriptions of hormone-based contraceptives. Providing teenage girls with access to hormone-based contraceptives, with reduced parental involvement, might be an effective way to reduce teen births.

Despite the effectiveness of SBHCs in reducing teen pregnancies, we find no evidence that they substantially reduce high school dropout rates. We measure high school dropout rates using reported high school diplomas awarded at the district level and U.S. Census and American Community Survey (ACS) data. Our estimates are universally small in magnitude, vary in sign across specifications, and are only rarely statistically significantly different from zero. Even for the largest of these point estimates, we can rule out at the 5% level that increasing primary care service hours equivalent to an average-sized SBHC would reduce high school dropout rates by more than 1.0 percent. The high school years might be too late in a child's life to substantially alter the likelihood of high school completion via improved access to primary health care.

Our most economically significant finding is that school-based health centers produce large declines in teen childbearing. There is much policy interest in reducing teen birth rates in the U.S. due to their high levels and the potentially high private and social costs associated with teen births (Kearney and Levine, 2012). At least for this outcome, these centers are quite effective at altering teen health. That they do not translate into changes in high school dropout rates underscores the importance of more research examining the role of health care services for school-age children in determining educational attainment.

3.2 School-based Health Centers

School-based health centers (SBHCs) are health clinics that are located inside specific schools or elsewhere on the school's property.[5] They are funded by various combinations of state and federal grants, in-kind donations by hospitals, donations from private foundations, and reimbursements from Medicaid and private insurance companies. School districts themselves typically do not provide direct financial support to SBHCs, other than providing space for them on school grounds. While SBHCs have been in existence since the 1930s, a surge in SBHC openings during the 1990's coincided with many states increasing revenues available to SBHCs using newly-available funds from tobacco company lawsuit settlements, cigarette taxes, and Maternal-and-Child-Health block grants from the federal government. Figure 1 shows the distribution of opening years for SBHCs in our data. Almost 83% of these SBHCs opened after 1989, with over 38% opening after 1997. Figure 2 shows the number of SBHCs in our data in each state relative to the size of the school-aged population in 2011. SBHCs are located in all but nine (mostly small) states. An eclectic mix of states such as Delaware, Louisiana, Maine, Maryland, New Mexico, Oregon, and West

[5]These are distinguished from community health centers that began opening in the mid-1960s to provided care to low-income communities as part of President Johnson's war on poverty. Bailey and Goodman-Bacon (2015) exploit the timing of the opening of these centers and show they had a significant effect on mortality rates of people over 50 years old. Relative to these centers, school-based health centers are focused on a much younger population with different health needs, and their prevalence is much more recent than general community health care centers. However, both types of centers are focused on bettering the provision of health care services to low-income communities.

Virginia have relatively large numbers of SBHCs per high school-aged child.

While cross-state variation in funding policies influenced the growth of SBHCs, our methodological approach is to exploit within-state variation in the timing of SBHC entry. SBHCs provide services for two main types of students: urban students in school districts serving low-income populations and rural students. As of school year 2010-2011, 54% of the centers were located in urban schools, with 28% located in rural schools and 18% in more suburban areas. Sixty-three percent of the students exposed to a school-based health center are of either African American or Hispanic descent.[6] Across similar communities in the same state, the provision of SBHCs may vary depending on relationships between school principals and local health administrators. While the specific requirements differ by state, typically it takes several steps to open the SBHC: 1) conduct a needs assessment to determine lack of access to health care among students, 2) build a partnership between a school and the local health organization (e.g., hospital, non-profit health clinic), 3) generate a funding plan, 4) find appropriate space in the school, 5) obtain approval from the state/local government, 6) develop a staffing plan that includes mechanisms for coordinating services across agencies, and 7) modify the space in the school so that it meets code for health clinics and has proper equipment. The impetus to open a center in a specific location can come from local health officials, school administrators, or community leaders. States typically require that an application for a new center is sent directly from the health organization that would operate the center, along with appropriate sign-off from the school district that would host the center. Given the bureaucratic and organizational hurdles associated with opening a center, it usually takes several years between initial conception and a center

[6]Online Appendix Table A-1 shows characteristics of counties and school districts with and without a SBHC by 2011.

opening. The unpredictability of both the location and timing of center openings provides the variation we need to estimate our models, and we conduct several tests to explore whether this variation is exogenous.

The focus of SBHCs is on providing primary care services for student populations. The majority of centers are attached to high schools, but many centers also provide services for students outside of the school to which they are attached: only 38% of centers report that use is restricted to students in the school. About a quarter of the SBHCs allow for families of the student to use the services, and 25% also allow use by school personnel. Almost 35% of the centers also report that they serve students from other schools. In some cases, the services provided are free to students. However, most centers operate more like traditional clinics and charge patients for services rendered. Due to the location of SBHCs, most students exposed by these centers are Medicaid-eligible, though, so these fees are unlikely to pose a large constraint to access. This feature of SBHCs highlights the fact that the treatment we examine is mostly due to health care *provision*, not due to health insurance access *per se*.

All centers provide primary care services, but the exact mix of services varies across centers. The distribution of primary care services is shown in Panel A of Figure 3. About 85% of centers also provide some form of reproductive health service. Panel B of Figure 3 shows the distribution of reproductive health services other than contraception provided by SBHCs in 2007-2008. Mostly, these services include testing for sexually transmitted infections, preventive care such as gynecological exams, PAP tests and prenatal care, as well as both abstinence and birth control counseling. Almost 40% of centers also are allowed to either prescribe or dispense contraceptives of some form directly, but many of the remainder refer students to other providers for contraception. Table 1 shows detailed information about the types of contraceptive services SBHCs offer. Over 37% either can dispense or prescribe the birth control pill, and another 30% can refer patients to other doctors for a prescription. Condoms are dispensed at over 30% of centers, and emergency contraception or plan B also is available either directly or through referral at the majority of SBHCs. Table 1 highlights that a large proportion of SBHCs provide significant contraceptive services but that there is considerable heterogeneity across centers in the types of contraceptives to which they give student access and the method by which students can access contraceptives. Because of the location of these centers, they may provide particularly important access to contraceptive services for female students who do not need to be taken to them by parents or guardians.[7]

In addition to primary care and reproductive health services, many school-based health centers have mental health and dental services. Eighty-four percent of centers provide oral health education, and 57% have dental screenings. Only about 20% conduct dental examinations, but the majority are able to refer students to dentists if they require dental services. Over 70% of health centers also have mental health providers on staff, with the remainder typically providing referrals through the primary care doctors for students who need mental health services.

Overall, SBHCs give students access to primary care doctors and nurses as well as more specialized medical services depending on the center. Since most centers can refer patients to more specialized doctors, the increased access to

[7]Currently, 26 states allow all minors over 12 to consent to birth control without their parents' approval. Another 20 states allow minors to consent under certain circumstances, such as being deemed "mature" or having a health issue. The remaining four states have no statutes regarding minor access to birth control.

primary care services that SBHCs represent is likely to increase health care options substantially for students who are served by these centers. The focus of this paper is on evaluating whether this increased access to health care affects teen birth rates and high school dropout rates. The main mechanisms through which these centers could impact student educational attainment are twofold. First, access to health care services could lead directly to better student health outcomes. To the extent that health enters positively in the production function for educational achievement, these health increases could drive better educational outcomes. A potential concern with this mechanism is that teens may be quite healthy. If high school students do not require much access to health care, then SBHCs will have little impact on them, at least in the short-run.

Despite the fact that high school corresponds with a relatively healthy part of the lifecycle, there is evidence that a substantial fraction of teens have health problems that would benefit from medical interventions. Figure 4 shows tabulations from the 2011 Youth Risk Behavior Surveillance System (YRBSS), which is a nationally-representative health survey conducted by the CDC that focuses on students in high school. As the figure demonstrates, the incidence of mental health issues and the prevalence of sexual activity amongst high school students is high. For example, almost 30% of students report feeling sad or hopeless, over 15% report considering suicide, and about 7% have attempted suicide. Almost 60% of these students have had sex, and many have done so without a condom or without any birth control. Furthermore, a non-trivial proportion of the sample reports being a victim of physical violence, and incidence rates of asthma and obesity are also high. Figure 4 shows racial/ethnic differences in these health outcomes as well, with black and Hispanic students reporting outcomes consistent with lower health levels and more risky behaviors. As discussed above, most health centers offer reproductive services that include birth control as well as pregnancy and STI testing. In addition, most offer mental health services. The tabulations in Figure 4 are suggestive that such services would be of value to many high school students.

There is further evidence of unmet health care needs among lower-SES high school students. In a review of the public health literature, Flores (2010) reports that the preponderance of work points to large disparities in adolescent health outcomes and health care access across the socioeconomic spectrum. Harris et al. (2006) show that about 25% of black and Hispanic adolescents report needing medical attention but not receiving it, as compared to about 18% for whites. About 7-10% of these adolescents also report being in poor health. Hence, there is ample evidence that teens in the U.S. have health outcomes and unmet health care needs that could lead SBHCs to have a substantial positive impact on their health and on their subsequent educational attainment.

Access to affordable primary health care can also reduce the household's exposure to financial risk from an adverse health event (Gross and Notowidigdo, 2011; Leininger, Levy and Schanzenbach, 2009; Finkelstein et al. 2012). Receipt of primary care services may make students healthier and allow them to address health problems before they worsen and cost more to treat. This effect of primary care service provision thus could better the financial position of households, which can lead to higher student academic attainment.[8]

Despite the rise in SBHC prevalence in the US over the past several decades, there is no nationally-representative study of these centers using methods that can plausibly identify their causal effects on health and education. Several prior

[8]See Michelmore (2013) and Dahl and Lochner (2012) for evidence on the effect of family income on student academic attainment.

analyses have examined the relationship between SBHCs and student health and educational achievement, and they typically show a positive relationship between SBHCs and these outcomes (Kerns et al. 2011; Walker et al., 2010; Geierstanger et al., 2004; Kisker and Brown, 1996). However, these studies have several serious shortcomings that we seek to address in this paper. First, all previous analyses have focused on identifying the effect of one SBHC or of several in a particular city or school district. No study of which we are aware has estimated SBHC impacts on health and academic outcomes for the entire United States. Results from the current literature thus are hard to generalize to larger state or national populations. Second, the previous work in this area largely has been cross-sectional in nature, either comparing outcomes across students who do and do not use the SBHC within a school or comparing student outcomes across schools with and without a health center. It is unlikely the set of control variables in the data sets used are sufficient to control for selection across schools or into SBHC use within a school. Thus, using cross-sectional methods in this context makes it very difficult to identify the causal effect of SBHCs on student educational attainment.

One recent study of SBHCs in New York City instead identifies the effects of SBHCs by examining longitudinal changes in academic performance among students who enrolled in elementary and middle schools shortly before and shortly after those schools added SBHCs (Reback and Cox 2016). New York is one of the only states in the country where SBHCs are restricted by law to only serve the students enrolled in the hosting school. They find evidence that the addition of SBHCs to elementary or middle schools increases students' scores on standardized tests in math and language arts. Their findings suggest that the health benefits from SBHCs could increase educational attainment, particularly if positive effects on middle school students do not fade during high school. Our work complements this analysis by examining the impacts of SBHCs serving high school students, by examining teen fertility, by providing both short term and longer term estimates, and by providing estimates for the entire US.

3.3 Data

The data for this analysis come from four sources: 1) National Alliance on School-based Health Care National Census of School-based Health Centers, 2) Live birth data from the U.S. Centers for Disease Control and Prevention National Vital Statistics System, 3) National Center for Education Statistics (NCES) data on high school diplomas awarded and enrollment, and 4) U.S. Census and American Community Survey data on school district dropout rates. Below, we discuss each of these data sources in turn.

3.3.1 NASBHC Census of School-based Health Centers

Beginning in fall 1998, the National Alliance on School-based Health Care began surveying school-based health centers about their locations, staffing levels, services provided, usage and the timing of when they first opened. They repeated their survey in fall 2001, 2004, 2007 and 2011. The survey is designed to be a census in the sense that all centers known to NASBHC are contacted, but there is considerable non-response. In the 1998 survey, 70% of centers contacted responded, and the response rates were 85%, 78%, 64% and 77% in 2001, 2004, 2007 and 2011 surveys, respectively.[9] Across all surveys, we observe 2,586 centers serving high school students in 566 school districts throughout the United States. This number of centers is larger than the total number of centers that exists in any one year, which is due to center closures over time.

Each NASBHC survey contains detailed information on center location (e.g., zip code), services, utilization, days and hours open, what populations the center serves, and staffing hours for both primary care and total medical staff. Primary care staff includes physicians and nurse practitioners only. Total medical staff hours include mental health, dental care, nurse and physicians' assistant hours in addition to primary care. Thus, for survey respondents, we have comprehensive information on the level and types of services the center provides for students.

We link centers over time across the different surveys to obtain a panel of SBHCs. The center identification codes NASBHC used changed over time, so that a unique id does not exist for each center. Instead, we match centers over time by linking them to the school districts in which they are located. Matching centers to school districts is complicated by the way centers report the schools that they serve. Since the survey question is open-ended, many centers give responses such as "all schools in district" or "only our schools" without naming the district or individual schools. Instead of relying directly on school names for the match, we use the geographic information about the center that was provided in the 1998, 2007 and 2011 waves. Centers in these waves were matched to school districts based either on their zip code or on their city and state. A

[9]Much of this non-response is actually due to center closures. Although NASBHC attempts to purge their roles of closed centers, which centers close is difficult to observe. Thus, the response rates among currently active centers is likely to be significantly higher than what is reported here.

school district was considered a match if it was the only district that shared this geographic information. Centers that could not be linked to school districts in this way, either because the geographic information applied to more than one district or the survey was missing information, were hand-matched to districts by using the NCES online school search tool. Centers were then matched to each other over time using the name of the center, the school in which the center is located, the schools the center serves, and the opening year. A center was matched across time if the name of the center and state were the same or the school location, name, and state were the same. Due to changes in reported names or school location, many centers had to be hand-matched across waves. It is important to highlight that the aggregation to the school district level means that errors made in matching specific centers to each other over time will not affect our results as long as we correctly link centers to school districts. Given the data limitations in the NASBHC data, using school-district level aggregations likely leads to less measurement error than if we had attempted to match each center to a specific school.

One of the drawbacks of our data is that we observe service and staffing levels only for the years in which the surveys were completed. However, for all but 51 centers (or 1.9% of the total centers observed), the opening date is contained in the survey.[10] These center opening dates allow us to use outcome data from before 1998. As Figure 1 demonstrates, 62% of the centers in our data were opened prior to 1998, so the use of these earlier data increases the amount of treatment variation considerably. For observations prior to 1998, we assume each SBHC has the service level equal to the first time we observe the center in the data. We linearly interpolate center service levels between surveys as well.

[10]We drop these 51 centers from our analysis, since we have no way of knowing when they first opened.

Furthermore, we assume a center closed when we no longer observe it in our data.[11]

3.3.2 Vital Statistics Birth Data

Data on all live births in the US come from the birth certificate files of the Centers for Disease Control and Prevention National Vital Statistics Data.[12] For each birth, we observe the race and ethnicity of the mother as well as her age. For mothers who live in counties with more than 100,000 residents, we also observe the county of birth. Recall from Section 2 that SBHCs are concentrated in urban and rural areas. The fact that geographic identifiers only are available for large counties means that our birth analysis is most relevant for the urban school-based health centers. The birth and SBHC data are merged based on the county of the SBHC. To the extent that school districts split county lines, we assign each center to the county in which it is located.

The vital statistics data give us information on all live births in 524 counties in the US from 1990 through 2012. Beginning the analysis in 1990 captures 86% of the SBHC opening variation in our data; we are loathe to extend the analysis sample back farther given that the first year we observe SBHC characteristics is in 1998. We construct "teen" birth rates – births per 1,000 women aged 15-18 – in each county and month.[13] To account for the timing differences between

[11]The way we identify center closings likely confounds closure and survey non-response for centers that respond to the survey in an earlier year but not subsequently. However, this method will bias our estimates towards zero to the extent that some centers we code as closing are still providing services to students. Furthermore, our instrumental variables strategy should account for any measurement error induced by center closures and non-response, as the instruments we use are unlikely to be related to closure or non-response.

[12]These data are available at http://www.cdc.gov/nchs/data_access/Vitalstatsonline.htm.

[13]For the remainder of the analysis, we refer to the birth rate among 15-18 year old women as

conception and birth, we link all births at the month-year level to the school year in which the conception took place assuming a 9 month gestation time. We then aggregate births to the school year-county level to construct a birth rate for each county and school year.

3.3.3 Common Core of Data High School Diploma Data

Since 1998, the National Center of Education Statistics has collected information on the number of high school diplomas awarded in each school district. These data are reported as part of the Common Core of Data (CCD).[14] We use these reports, combined with grade-specific enrollments, to construct a measure of high school dropout rates. Specifically, we estimate the dropout rate for a given grade as $1-\frac{Diplomas_t}{Enrollment_{t-g}}$, where g$\in$ $[0,1,2]$. For example, when g=2, this formula yields the 10th grade dropout rate. In particular, it is the proportion of 10th graders in the district from two years ago that do not receive a high school diploma this year. Similarly, we calculate the 11th and 12th grade dropout rate using once-lagged enrollment of 11th graders and year t enrollment of 12th graders. We calculate these rates for each school district in the US, from 1998-2010.[15]

Heckman and LaFontaine (2010) and Mishel and Roy (2006) provide detailed discussions of the problems arising from using the CCD diploma data to calculate graduation rates.[16] The biggest problem with these data is associated with the use of 9^{th} grade enrollments, as there is a substantial amount of grade retention in 9^{th} grade. This grade retention is more prevalent for low-SES students

the "teen birth rate."

[14]The CCD diploma data are available at http://nces.ed.gov/ccd/drpagency.asp.

[15]Because diploma data are from the spring of each year and the SBHC surveys are in the fall, we lag all graduation rates by one year to align them with the SBHC service data.

[16]See also the comprehensive review of U.S. high school graduation rates in Murnane (2013).

as well, and it leads one to understate graduation rates, especially for minority students. Heckman and LaFontaine (2010) show that when one uses 8^{th} grade enrollments instead, this bias is reduced considerably. We instead ignore 9^{th} grade enrollment and focus on enrollment in higher grades that are less problematic. To the extent that SBHCs affect the likelihood of being held back in 9^{th} grade, we thus will miss some of the ways in which these centers influence students' paths through high school. However, our estimate should not be seriously affected by the retention rate problems that come with using 9^{th} grade enrollment data.

The CCD diploma data cannot distinguish between actual dropout rate changes and changes in the timing of degree receipt and student transferring behavior. Thus, this dropout rate will predominantly measure "on time" high school graduation for those in each grade cohort net of transfer. If there is a net loss of the 10^{th}-12^{th} grade cohorts due to transferring out of the school district, however, this measure will show an increase in dropout rates. For transferring to create a bias in our estimates, it would have to be correlated with SBHC entry/exit and service changes. While possible, we do not believe such effects would be large. The complications induced by these data are balanced by the fact that they are yearly, allowing us to exploit more within-district variation in SBHC services. Table 2 presents descriptive statistics of dropout rates calculated using these data.

3.3.4 US Census and ACS Data

We supplement our graduation analysis with 1990 and 2000 Census data as well as with the 2005-2011 American Community Survey. Using these data, we calculate for each school district the proportion of 14-17 year olds living in the school district who are not enrolled in school and who do not have a high school degree. This is the 14-17 year old dropout rate. The 18-19 dropout rate is calculated similarly using those aged 18-19. These data provide several advantages over the diploma data. First, they allow us to distinguish between males and females. Given our focus on teen fertility rates and the fact that males are more at risk of dropping out, it is useful to examine dropout effects by gender. Second, high school degrees in the Census/ACS include GEDs while the diploma data do not. Even though the returns to a GED are lower than the returns to a traditional high school diploma (Heckman and LaFontaine 2006), it is important to distinguish between any shifts across degree types versus any change in overall degree attainment. To the extent that the Census/ACS and CCD graduation rate estimates yield similar results, it suggests that our estimates are not being driven by changes in the proportion of students receiving a GED. A drawback of these data is that we only observe each school district a maximum of 4 times: in 1990, 2000, 2005-2007 and 2008-2011. But, combined with the diploma results, this analysis provides a more complete picture of the effect of SBHCs on high school completion. Because the ACS data are for a period of 3 years, we use the average SBHC service level over those 3 years for each school district when we analyze these data. Descriptive statistics of the dropout rates in the Census and ACS are shown in Table 2.

3.4 Empirical Methodology

Our methodological approach to overcoming the inherent endogeneity between health care access, health and educational attainment is to use the variation in student exposure to health care services that is driven by school-based health center openings and the scope of the services provided. Our baseline model is a straightforward difference-in-differences design that uses variation only from the initial center entry in an area. We compare changes in birth or graduation outcomes in areas that receive their first center relative to areas that do not receive their first center in that year. Due to data limitations, our birth rate analysis and completion rate analysis occur at different levels of aggregation. In the birth data, the county is the most disaggregated level of geography available, so this part of the analysis is done at the county level.[17] In particular, we estimate models of the following form:

$$Y_{cst} = \beta_0 + \beta_1 SBHC_{ct} + \gamma_c + \delta_{st} + \epsilon_{cst}, \tag{3.1}$$

where Y_{cst} is the birth rate per thousand women aged 15-18 in county c in year t,[18] γ is a set of county fixed effects, and δ is a set of state-by-year fixed effects that control for any state-level unobserved shocks in each year as well as state-year level policies (such as Medicaid). The variable $SBHC$ is an indicator variable equal to 1 if there is a school-based health center in the county and is zero otherwise. Thus, the variable of interest in equation (1) is β_1, which shows the effect on the birth rate of a SBHC entering the county.

[17]One benefit of using aggregated data is that our estimates account for both the direct effect of SBHCs on teen pregnancy and the indirect effects coming through peer influences that Yakusheva and Fletcher (2015) show are important.

[18]We also have estimated models that use the log of the birth rate. These estimates are very similar to those shown below and available from the authors upon request.

The county fixed effects control for any fixed differences across counties in birth rates that are correlated with SBHC treatment. The identifying variation for β_1 comes only from differences in the timing of the first center opening across counties. Identification of β_1 thus rests on several assumptions that are common in difference-in-differences analyses. The first is that the decision to open a center is uncorrelated with trends in teen birth rates. Put differently, counties in which a SBHC will open in the near future should have the same outcome trends as those that will not experience an initial opening in the near future. Of particular concern is whether centers are put into schools where the teen birth rate is declining. If so, equation (1) will not be able to distinguish treatment effects from differential secular relative trends. We do not believe, however, that this concern is very relevant in this context. It is far more likely that SBHCs are targeted toward schools that have declining health outcomes. As discussed in Section 2, the timing of when centers open is likely to be related to lack of health care access among students, the desire and ability of a principal or administrator to partner with a local health care provider, space in the school, and demand among the community for expanding health care access for low-income kids. Many of these factors may be related to underlying trends in health or educational attainment, but the sign of any resulting bias would be towards zero.[19] We test directly for whether center entry is related to pre-SBHC birth rate trends with the following "event study" specification:

$$Y_{cst} \quad = \quad \phi + \sum_{\tau=\leq -6}^{\geq 11} \alpha_\tau I(t - t_0 = \tau)_{ct} + \gamma_c + \delta_{st} + \epsilon_{cst}. \qquad (3.2)$$

In equation (2), $I(t - t_0 = \tau)$ is an indicator variable equal to 1 if the observation

[19]We also note that if the timing of center openings were related to unobserved trends, our birth rate and dropout rate estimates should be biased in the same direction. That we find no effect on high school dropout rates but a large negative effect on teen fertility rates argues against such selection.

is τ years away from the first SBHC opening in the county and is equal to zero otherwise. These variables are zero for counties that have no health centers in the time period of our analysis. This event study model allows us to both test for pre-treatment trends by examining $\alpha_{-5} - \alpha_{-1}$ and to test for time-varying treatment effects (given by $\alpha_0 - \alpha_{10}$) that might be missed in equation (1). We focus on the event window from relative year -5 to 10 as outside that window we have fewer observations with which to identify each relative time parameter. We group together observations with event time less than -5 and observations with event time greater than 10 in order to avoid altering the analysis sample. The model includes all "never-treated" counties as well, which constitute the implicit control group.

Another identification concern with difference-in-differences analyses is that secular shocks or unobserved policies that correlate with the timing of the treatment can bias the results. Such shocks are unlikely to be a factor in this analysis, however. Since the timing of the treatment varies across counties, it is doubtful secular shocks exist that are highly correlated with the timing of SBHC entry. That it takes several years for centers to open from when they are initially conceived also makes it unlikely that they are systematically correlated with county-level shocks. As well, the use of state-by-year fixed effects helps control for any state-level policies or shocks that could be correlated with the timing of treatment. Nonetheless, it could be the case that policies disproportionately impacting low-income communities (such as welfare reform and EITC expansions) are passed in similar time periods to when centers entered. In Section 5.1.2, we show our estimates are robust to controlling for separate time effects for low-income counties and to allowing for different state-year fixed effects by whether the county median income in 1990 is below that of the median county in the state. We also conduct falsification tests using birth rates among women in their twenties that confirm that the birth effects are isolated to high-school-aged women. The results of these falsification tests are inconsistent with the idea that important alternative policies or secular trends were correlated with the rollout of SBHCs.

The coefficient β_1 in equation (1) yields the average effect of center entry. This treatment specification omits a large amount of heterogeneity across centers, though, in the amount and type of services offered. From a policy perspective, we are interested more in the services offered through the centers than the centers *per se*. We therefore estimate versions of equation (1) that replace $SBHC_{ct}$ with *Service Hours*$_{ct}$, which are measures of the amount of services provided by each center relative to the underlying size of the student population. Specifically, we focus on two different service measures: Primary Care Staff Hours per week and Total Medical Staff Hours per week. These services are set to zero prior to an SBHC opening. The Total Medical Staff Hours differ from Primary Care Hours due to hours from mental health staff, dental staff, physician's assistants and nurses. In the Online Appendix Table A-2, we also show estimates that use Days per Week or Hours per Week as the service measures. As Primary Care Staff and Total Medical Staff Hours are the most comprehensive measures of the medical services provided by school-based health centers, they are our preferred treatment variables. Means of these treatment measures are shown in Table 2.

Throughout the analysis, the SBHC service variables are constructed by first summing the total amount of each service measure for each county or school district and year. For example, we calculate the total number of medical staff service hours in the county and year across all centers in the county. We then divide by the total high-school-aged population in the county.[20] This provides a measure of the hours of SBHC medical services per high-school-aged student in the county. Finally, we re-scale the measure to be representative of a typical center by multiplying by 1000, which is the approximate average size of a high school in our sample. The method is identical for our school district level regressions, where we sum over districts rather than counties. Using the primary care or medical staff hours as our treatment measures, β_1 is interpreted as the effect on the birth rate of SBHCs increasing their service levels by an additional hour. When multiplied by the average SBHC service level, this estimate shows the effect of a service increase equivalent to one more average-sized center opening. We focus on this parameter for policy purposes.

Variation in primary care and medical staff hours comes from two different sources: 1) openings/closings of SBHCs with different service levels and 2) changes in service levels among open centers from year to year. In addition to the identification assumptions discussed above, we now require that decisions about the amount of services each center offers are uncorrelated with pre-treatment trends in teen birth rates. If service levels rise in areas that were already beginning to experience rising or falling teen birth rates, then our estimates of β_1 will be biased.

We address these core identification concerns in several ways. First, we estimate event study models that test for selection on trends as a function of initial service level variation:

$$Y_{cst} = \phi + \sum_{\tau=\leq -6}^{\geq 11} \alpha_\tau \textit{Service Hours}_{ct_0} * I(t - t_0 = \tau) + \gamma_c + \delta_{st} + \epsilon_{cst}. \quad (3.3)$$

[20]Our high-school-aged population count includes individuals between the ages of 15 and 19.

The *Service Hours* variable in equation (3) is set to the first observed service level in that county. That is, we set it equal to the service level observed when τ=0, denoted t_0. This model thus tests for selection related to initial service levels as well as time-varying treatment effects by initial service levels. We also estimate a version of this model in which we allow *Service Hours* to vary over time after initial entry, similar to how it is specified in equation (1). Comparing the post-entry estimates across these two versions of this model provides evidence on whether post-entry variation in service levels is exogenous.

Second, in order to account for the potential endogeneity of year-to-year service level variation, we employ an instrumental variables strategy. We instrument Primary Care Staff Hours and Total Medical Staff Hours with an indicator for whether there is a center in the county and with a quadratic trend in the time since first center entry. This quadratic time trend is set to zero prior to the first center entering a county. As long as center entry is uncorrelated with pre-entry trends in birth rates, this instrument is valid. Thus, equation (2) is a direct test of the validity of this IV approach. Critically, because the instruments do not contain variation in service levels after the initial center is opened, they will account for any endogeneity in equation (1) in year-to-year service levels.

Our analysis of high school dropout rates takes a very similar form as our birth rate models. The main difference between the two is that, for high school dropout rates, we observe outcomes at the school district level, rather than at the county level. We estimate the following models:

$$
\begin{aligned}
Y_{dst} &= \beta_0 + \beta_1 SBHC_{dt} + \gamma_d + \delta_{st} + \epsilon_{dst} && (3.4)\\
Y_{dst} &= \phi + \sum_{\tau=\leq -6}^{\geq 11} \alpha_\tau I(t-t_0=\tau)_{dt} + \gamma_d + \delta_{st} + \epsilon_{dst} && (3.5)\\
Y_{dst} &= \phi + \sum_{\tau=\leq -6}^{\geq 11} \alpha_\tau \mathit{Service\ Hours}_{dt_0} * I(t-t_0=\tau) + \gamma_d + \delta_{st} + \epsilon_{dst}. && (3.6)
\end{aligned}
$$

In equations (4)-(6), we now include district, rather than county, fixed effects. The assumptions underlying the identification of the treatment parameters in equations (4)-(6) are essentially identical to those for equations (1)-(3), except instead of there being no differential county-level relative trends, here there must be no differential district-level trends. Equations (5) and (6) allow us to test for such trends as well as for time-varying treatment effects. We also estimate instrumental variables models akin to those at the county level to account for any endogeneity associated with yearly variation in SBHC service levels in our dropout analysis.

A final potential methodological issue is the presence of measurement error in our service hours treatment measures. One source of measurement error is the fact that, while the NASBHC National Census is designed to cover all health centers, there is not complete coverage in every year. The use of multiple years of data combined with information on the date of opening of the centers should mitigate this problem. But, it is possible there are health centers we do not observe in our data and some we code as closing when they still exist. To the extent that some districts and counties are more heavily treated than our data show, this should attenuate our OLS estimates. The instrumental variables model estimates should avoid similar attenuation, however, because the instruments are unlikely to be correlated with center closure or with survey non-response.

A second source of measurement error is that prior to 1998, the first year of NASBHC data, we cannot observe changes in the level of services provided. For all centers opened before 1998, we use the first observed service levels (typically from the 1998 survey). This could produce further measurement error in the *Service Hours* variables. Finally, aggregation to the county and school district levels could produce measurement error because many students in each county and district do not have centers in their own school buildings. Some aggregation would be appropriate even if it were not necessitated by the data, because 62% of centers are open not only for students in the hosting schools but also for other community residents. Furthermore, SBHCs are concentrated amongst the lowest-SES schools in counties and districts, which also are schools in which teen pregnancy and dropout rates are most prevalent.[21] This argument supports our contention that the aggregated data can provide informative estimates of the relationship between school-based health centers, teen childbearing, and educational attainment.

[21]We also note that it would be exceedingly difficult to match schools to specific centers. The school codes for centers are not consistently present in the data, and many centers have administrative offices that occasionally answer the surveys. In some years the administrative offices answer the surveys and in some years the centers themselves do. Aggregating to service levels at higher geographic levels sidesteps this problem.

3.5 Results

3.5.1 Birth Results

Main Estimates

Table 3 presents the baseline estimates of the effect of school-based health centers on teen birth rates. Each cell in the table is from a separate regression, and all standard errors are clustered at the county level. In the first column, we show OLS estimates of β_1 from equation (1). The top row shows results using an indicator for the presence of any center in a county as the treatment measure. When the first SBHC enters a county, the teen birth rate declines by 1.3 per 1000, which represents a 3.0% percent decline. The remaining rows show estimates using Primary Care Staff Hours and Medical Staff Hours as the treatment measures. Across these two treatment variables, the table shows a consistent negative relationship between SBHC service levels and teen birth rates. Ten additional primary care staff hours or medical staff hours per week decreases teen births by 2.18 or 1.08 per 1,000 respectively. A useful way to interpret these estimates is to calculate their implications for the effect of opening an average-sized center. To calculate such an effect, we multiply the estimates by the average amount of services each center supplies (shown in Table 2) and then divide by the average birth rate for this group. The estimates suggest that adding an average-sized center in a county would reduce birth rates by 2.4 or 2.7%. The magnitude of these estimates is similar to, if somewhat smaller than, the 6.8% decline in birth rates among 18-19 year olds following Medicaid family planning waiver expansions reported in Kearney and Levine (2009).

A central concern with the type of difference-in-differences analysis we employ is that centers may be targeted at areas based on preexisting trends. Figure 5 shows the estimates of α from equation (2). We have excluded relative year -1 such that all estimates are relative to this year. The points in the figure show the point estimates of α, while the lines extending from each point show bounds of the 95% confidence intervals that we calculated using standard errors that are clustered at the county level. There is no evidence of negative pre-treatment trends in birth rates. The pre-treatment trend line is flat, especially within 5 years of center entry, and the pre-treatment coefficients are jointly insignificant (p-value of 0.56). These results suggest that center entry is exogenous with respect to teen birth rate trends.

The second pattern evident in Figure 5 is that the long-run effects of SBHCs are much larger than the short-run effect. These are at least three potential explanations for these rising effects. First, repeated exposure during all four years of high school should produce larger effects for students than exposure for a smaller number of years. Second, centers may take time to ingrain themselves in the community. Third, many counties initially opening a center later expand their services and have subsequent center openings. Controlling for state-by-year effects, we find that county-level service hours hit their peak 5 years after the first center opens and district-level service hours hit their peak 8 years after the first center opens.[22]

While the timing of center openings is exogenous with respect to pre-existing teen fertility rate trends, the amount of services they offer are correlated with

[22] These tabulations come from event study analyses of how service levels vary after initial center entry. Online Appendix Table A-6 shows similar estimates that impose a quadratic time trend. While service levels grow modestly in the several years following initial center entry, the majority of the service level variation is driven by the timing of first entry.

these trends. Both initial service levels and post-entry service level variation appear to be related to pre-entry fertility trends (as displayed in Online Appendix Figure A-1).[23] More services appear to be targeted towards areas experiencing increasing teen fertility rates, which attenuates the OLS estimates that use service hours as the treatment measure. To account for the endogeneity of service levels, we instrument for Primary Care Staff Hours or Medical Staff Hours with the timing of the first center opening. Our instrumental variables are an indicator variable for whether a center has opened and quadratic time trends for the number of years since that first center opened. The results from these IV models are shown in column (ii) of Table 3. The instruments are strong, with first-stage F-statistics above 50.[24] The estimates in column (ii) are considerably larger in absolute value than the OLS estimates and are statistically significantly different from zero at the 5% level. The larger estimates are due to the fact that the instruments account for the targeting of services based on teen fertility rate trends. In our IV models, opening an average-sized center reduces teen birth rates by over 5%. On the whole, the results in Table 3 tell a consistent story that opening an SBHC in a county has a sizable negative effect on teen birth rates on the order of 3-5 percent.

At which age are teens' fertility rates most affected by SBCHs? In theory, SBHCs may affect both younger and older teens. On the younger end, middle school students might also be able to visit centers. On the older end, the impact of centers on sexual behavior and use of contraception may persist beyond

[23]Panels A and B of Figure A-1 presents evidence that counties with higher initial hours of service have rising birth rates, although the estimates are not jointly significant. The increase in birth rates continues for the first year after the center has opened. In Panels C and D, we hold service hours constant at their initial levels both pre- and post-initial entry and show a steady decline in birth rates after the first center opens. Taken together, these results suggests that the counties continuing to experience relatively high birth rates tend to increase their centers' service hours the most.

[24]First-stage coefficients are shown in Online Appendix Table A-6.

a woman's time in high school. Table 4 presents estimates by age, including 19 year olds and those under 15. For each age, we show estimates using the same models and treatment measures as shown in Table 3. A consistent pattern of results emerges in Table 4: health care services from SBHCs reduce teen birth rates among teens of all ages, with the largest proportional effects coming from the youngest teens. For example, in our 2SLS models predicting primary care hours, an average-sized SBHC reduces births among girls 14 and under by 15.2%, among 15-year olds by 12.0% and among 16-year olds by 9.6%. Among teens aged 17 or 18, the estimated effects are less than 6%. Across model specifications, the estimates for 19 year olds are much smaller and often are not statistically significant; this is sensible, because many of these women are no longer enrolled in high school and may thus have far less access to SBHCs. Proportionally larger effects for younger women is an important finding, because the private and social costs of teen fertility may be highest for the youngest mothers. That SBHCs have such a large effect on young teen births suggests they are most successful at reducing fertility among the population that is of highest concern among policymakers.

SBHCs differ in the types of contraceptive services they offer. About 65% of centers offer some type of birth control, either directly or through referral (see Table 1). In Table 5, we show estimates of equation (1) that allow the effect of Primary Care Staff Hours and Medical Staff Hours to differ by the type of contraceptive services offered by the clinic. We split centers into four groups that together encompass the entire range of birth control offerings in US SBHCs: centers that prescribe hormone-based contraceptives on-site,[25] centers that re-

[25]Hormone-based contraceptives include birth control pills, Depo-Provera, implants, interuterine devices (IUDs), the patch, and the NuvaRing. We code centers as offering hormone-based contraceptives if they report offering birth control pills or report offering more than one other form of hormone-based contraception.

fer patients for hormone-based birth control but do not offer condoms on-site, centers that refer patients for hormone-based birth control and offer condoms on-site, and centers that do not offer any contraceptive services. Each column in Table 5 comes from a separate regression. The results are broadly consistent with SBHC services most affecting teen birth rates in centers that can prescribe hormone-based birth control, but the estimates are somewhat imprecise and the differences in slopes are not statistically significant at conventional levels. As shown in Figure 3, many of the centers that do not offer contraception do offer other family planning services, such as pregnancy tests, tests for sexually transmitted infections, abstinence counseling, and general health advice that would come with a primary care visit. Table 5 reveals that our results are not driven by condom distribution, which is consistent with theoretical and empirical research arguing that distributing condoms may not reduce (and might increase) teen birth rates (Buckles and Hungerman 2014; Arcidiacono, Khwaja and Ouyang 2012). The results suggest that providing female teenages with easier access to hormone-based contraception, access that does not require them to go through their parents, may substantially decrease teen fertility rates. SBHCs might also reduce rates of sexually transmitted diseases.[26]

School-based health centers may have a larger effect on African American and Hispanic students than on white students because these centers are targeted

[26]We estimated state-level models of how SBHC services affect STD rates among teens using data from the U.S. Center for Disease Control, which are shown in Online Appendix Table A-5. Unlike birth data, data by age group for STDs are available at the state level and not the county level (county level data are not disaggregated by age). We regressed rates of three STDS - gonorrhea, chlamydia, and syphilis – among 15-19 year olds on the number of hours of primary care and total medical staff services provided by school-based health centers in that state, in models controlling for state fixed effects and year effects. Although most of the estimates are not statistically different from zero, they all point to sizable declines in STD rates among teens when SBHC services in the state rise. While the need to aggregate to the state level leaves us with too little power to draw definitive conclusions, these results are suggestive of positive sexual health benefits of SBHCs in addition to lower teen birth rates.

at low-income populations. Furthermore, African American and Hispanic teen birth rates are much higher than those of whites, which makes these groups particularly important to study. In Table 6, we show OLS and IV estimates of the effect of SBHCs on teen fertility rates by race/ethnicity. As in Table 3, we also calculate percent effects of opening an average-sized center in order to compare more easily across specifications. There is no evidence that opening a SBHC reduces teen birth rates among whites. The estimates are universally small and are not statistically different from zero. This non-result is likely driven by the fact that our treatment is at the county level, and white students in treated counties are far more likely to be in wealthier areas that do not have a center. Thus, within a treated county, whites are less likely to be actually exposed to a center than black and Hispanic students. As a result, SBHCs have a much larger impact on birth rates among black and Hispanic teens. In the 2SLS models, adding an average-sized SBHC reduces both black and Hispanic teen birth rates by about 8%. Although the Hispanic estimates are imprecise, these results demonstrate that school health centers affect teen births most among racial and ethnic minorities who are more likely to live in low income areas that have such centers.

Robustness Checks

As discussed in Section 4, one of the central identification assumptions underlying our approach is that there are no secular trends or shocks that align with the rollout of SBHCs across areas. Of particular concern is whether federal or state governments passed policies disproportionately affecting cities with higher concentrations of low-income residents during the same time period when many SBHCs opened. Many centers opened in the mid-1990s (Figure 1), a time period in which welfare programs were reformed, the EITC was expanded, and many states were expanding public health insurance programs. The state-year fixed effects should account for these policy changes if they affected teen birth rates in all counties similarly in a state and year. But these policies disproportionately affected low-income communities, so they may have disproportionately affected teen birth rates in these communities. Table 7 shows several robustness checks to compare with the results from Table 3. Columns (i) and (ii) of Table 7 reveal that the results are robust to controlling for state-year-median income fixed effects by allowing the state-year fixed effects to differ based on whether a given county's median income in 1990 was in the bottom half of all counties in that state. Columns (iii) and (iv) of Table 7 show that the results are also robust to controlling for both state-year fixed effects and differential year fixed effects among the bottom 20% of counties in the US according to 1990 median earnings. The estimates in columns (i) through (iv) are similar to those in Table 3, with somewhat larger estimates for the IV models. Table 7 thus suggests that our main results are not upwardly biased by state or national policies aimed at lower-income communities.

Next, we relax the linear functional form assumption between SBHC service levels and outcomes. We do this by controlling separately for SBHC services and for the existence of a center in the county rather than just the interaction of these two measures. This is a more flexible way to model the treatment, and the estimates in column (v) of Table 7 show that this leads to a slightly larger percentage effect of SBHCs on teen fertility. If anything, the functional form embedded in our baseline estimates leads to somewhat conservative estimates.

Another check for the existence of secular trends or shocks that can bias our estimates is to examine whether the short run effects of SBHCs are limited to teen women. If older women also experience declines in birth rates when SBHCs enter, this could be evidence that the emergence of these centers is correlated with other factors affecting fertility rates for women of all ages. This falsification test is complicated by the fact that many older women were treated by SBHCs when they were younger and by the fact that many centers are open to the community at large. We examine birth rates among women aged 20-24 and aged 25-29, and we restrict the sample to counties that did not have a center when women in these age ranges were of high school age. We also restrict our sample to states in which fewer than half of SBHCs report that they serve non-students. Table 8 shows these results; for both age groups, there is no evidence of a decline in births associated with SBHC entry. Indeed, birth rates among 25-29 year olds increased slightly in counties with greater intensity of services from SBHCs. These results are not simply due to the change in sample: column (iii) repeats the analysis on the same sample for 15-18 year olds. The results in column (iii) are similar to those shown in column (i) of Table 3. That the fertility effects of SBHCs are isolated to those who are of high school age strongly supports our identification strategy.

3.5.2 High School Dropout Results

The results presented above suggest that school-based clinics promote better health outcomes among the teens exposed to them, at least in terms of birth rates. A question of high importance is whether the changes in teen health caused by these centers, in terms of pregnancy as well as other health outcomes, affect educational attainment. For students in the low-income areas targeted by SBHCs, high school completion is a very important measure of educational attainment, and it thus is the focus of our analysis. In Table 9, we present the first evidence in the literature on the effect of providing primary care services to low-income school-age children on high school dropout rates. Due to serial correlation of errors within districts over time, all estimates are accompanied by standard errors that are clustered at the school district level throughout the dropout rate analysis.

The estimates in Table 9 are in percent terms, such that a coefficient of 1 would mean that a 1 hour increase in SBHC services would increase dropout rates by 1 percent (rather than by 100% if the dependent variable was in percentage terms). Across all models and treatment measures, there is little evidence that SBHCs or SBHC services affect high school dropout rates.[27] Roughly half of the estimates are positive, and only one of the estimates is statistically significant at even the 10% level. Furthermore, the estimates are precise: the 95% confidence intervals show we can rule out declines in dropout rates from an average-sized center of more than -0.5% for 10^{th} grade, -1.0% for 11^{th} grade, and -0.7% for 12^{th} grade.[28]

Figure 6 shows the event study estimates from equation (5) for 10^{th}, 11^{th} and 12^{th} grade dropout rates. Estimates of equation (6) using Medical Staff Hours and Primary Care Staff Hours are shown in Online Appendix Figures A-2 through A-4. These figures show that the null finding in Table 9 does not

[27] See Online Appendix Table A-6 for first-stage IV estimates.

[28] Similar to Table 5, we have estimated dropout models that examine heterogeneous SBHC effects by birth control services offered. These are shown in Online Appendix Table A-3 and do not point to any dropout rate effects in centers that offer access to certain types of contraception. We also have examined effects of service hours among centers that offer mental health services. We find no evidence of a dropout rate effect among centers that offer such services. These results are available from the authors upon request.

mask important heterogeneity in long-run effects or selection on pre-treatment trends. Recall that our dropout rate sample begins in 1998, and as a result we have much fewer observations pre-dating center openings. Thus, the standard error bounds in the pre-treatment period are relatively large. Still, there is little evidence of differential trends prior to center entry, and there is no evidence of a dropout effect post-entry either in the short or long run.

Dropout rate estimates using Census/ACS data are shown in Table 10.[29] Similar to Table 9, we fail to see statistically significant effects of SBHCs on high school dropout rates. The one exception is for women aged 18-19. When we use Primary Care and Medical Staff Hours as the treatment measure, there are small, negative effects of an average-sized SBHC on the dropout rate. However, these estimated effects are no more than one quarter of a percent, are not robust to using a center indicator as the treatment measure, and are not statistically significant at the 5% level. Thus, we view these estimates as being consistent with at most a very small impact of SBHCs on female high school dropout rates.

Our findings relate to a large literature examining the causal effect of teen childbearing on educational outcomes. While there is a robust positive correlation in most data sets between teen pregnancy and the likelihood of dropping out of high school, obtaining credible causal evidence of this link has proven difficult. The difficulty in establishing causality in this context is that it is very hard to generate variation in teen pregnancy rates that is driven by factors that do not affect schooling decisions as well. The literature on this subject, while large, is quite mixed. Ribar (1994) uses age at menarche, OB-GYN availability and state abortion rates as instruments and finds no effect of teen childbearing

[29] The limited number of observations per district preclude us from estimating IV models with these data.

on high school completion. Hotz, McElroy and Sanders (2005) use natural experiments driven by miscarriages to generate plausibly exogenous variation in teen births. They find a small negative effect of teen childbearing on high school completion. Fletcher and Wolfe (2009) and Ashcraft, Fernandez-Val and Lang (2013), however, argue that miscarriages are not exogenous events; they report modest negative effects of adjusted teen birth effects on high school completion. More closely related to this study, Klepinger, Lundberg and Plotnick (1999) use state-level variation in family planning and abortion services/policies as instruments for teen childbearing. They report that a teen giving birth reduces her educational attainment by 2.5 years. Finally, there are several studies that use sibling fixed effects as well as matching estimators to identify the effect of teen childbearing. While the sibling fixed effects analyses come to very mixed conclusions (Ribar, 1999; Holmlund, 2005; Geronimus and Korenman, 1992), the results from the matching literature point more consistently to a negative effect of teen fertility on educational outcomes (Levine and Painter, 2003; Sanders, Smith and Zhang, 2008). Our estimates, however, suggest that the teen birth rate declines as well as the other health benefits teens receive as a result of these centers do not substantially affect their likelihood of completing high school.

One explanation for the lack of an effect on dropout rates in the presence of a teen birth effect could be that the birth estimates only use data from large counties. To explore this potential explanation, in Online Appendix Table A-4 we estimate dropout rate models using the CCD diploma data in which we use only the counties included in the birth rate analysis. The results are extremely similar to baseline and show no effect of SBHC services on high school completion. Thus, the difference between the birth and dropout findings is not due to the differences in the samples used.[30]

Another alternative explanation for the lack of a dropout rate effect is that SBHCs lead to a reduction in school resources that counteract any health effects. These centers are not financed by the school, and they do not use school resources aside from the space that they are allocated. However, it still is possible that SBHCs use other school resources in a manner that might influence our dropout rate estimates, or SBHC entry could be correlated with unobserved trends in school resources. In Table 11, we examine whether SBHC service variation is correlated with school expenditures using data from the 1998-2011 Common Core of Data. We see that there is no relationship between SBHC services and per-student expenditures: the coefficients are small, precisely estimated and are not statistically different from zero at even the 10% level. These results suggest that there are no expenditure changes correlated with SBHC entry or service level variation that could bias the results and conclusions of our analysis.

3.6 Conclusion

Disparities in health care access, health and educational attainment are large in the United States, and policies to help close these gaps have received much policy attention. In this paper, we study school-based health centers that provide primary health care services to students and families living in under-served communities. Despite the rapid growth of SBHCs in the US over the past two

[30]In results available upon request, we also have estimated dropout rate models aggregated to the county-year level rather than the district-year level. The estimates are very similar to those shown in Table 9.

decades, the effect of these centers on health and educational attainment has not been studied previously in a manner that allows one to overcome the endogeneity problems related to center placement and use decisions. Using detailed data from repeated surveys of SBHCs conducted by the National Alliance on School-based Health Care, we construct district- and county-level measures of SBHC services over time and employ difference-in-differences and instrumental variables techniques to identify the causal effect of these center services on teen fertility rates and on high school dropout rates.

We present two broad findings from our empirical analysis. First, we show the SBHCs have negative effects on fertility rates among teenage girls. Adding a center with the average amount of SBHC services leads to a decrease in the 15-18 year old birth rate of about 5% relative to the baseline fertility rate. These effects are larger for younger teens, and they are concentrated among African American and Hispanic teens who are most likely to be exposed to a center. Second, despite the large effect of SBHCs on teen fertility, we find no substantial effect on high school dropout rates.

There are several implications of our results that are important for public policy. One central message of our findings is that SBHCs are a useful tool to reduce teen birth rates in the US, which are among the highest in the industrialized world (Kearney and Levine 2012). Another important implication of our results is that the provision of low-cost and convenient primary care services through schools has at most a small effect on students' decisions to drop out of high school. This is not to suggest that providing such services does not improve these students' lives, but it does suggest that any positive health benefits of this care do not immediately yield greater educational investment.

High school health interventions may come too late to influence high school completion; it is possible that expanding health care services to these children when they were younger would have produced greater effects on high school completion rates. Our work highlights the importance of further study of the linkages between health care access, health outcomes and educational investment decisions to determine whether there are aspects of health care provision that could support educational investment among students from low-income backgrounds.

3.7 Tables

Table 3.1: Percent of Health Centers Providing Different Contraceptive Services

Contraception Type	Prescribed & Dispensed On Site	Prescribed On Site	Referrals Only	No Provision
Condoms	31.7	N/R	N/R	68.3
Birth Control Pills	22.1	15.5	29.9	32.5
Birth Control Shot (Depo-Provera)	25.7	7.7	32.1	34.6
Implant	4.9	6.6	47.1	41.5
IUD	4.3	6.6	49.6	39.5
Patch	14.6	12.5	34.9	38.0
Ring (NuvaRing)	16.5	12.2	33.5	37.9
Emergency Contraception	20.0	11.0	30.2	38.8
Any Hormone Contraception	26.9	9.2	28.2	35.7
Any Contraception	31.7	6.6	26.4	35.3

Source: 2011 National Alliance on School-based Health Care census data. Hormone contraception includes all listed methods except condoms and emergency contraception. "N/R"=not relevant for that category.

Table 3.2: Descriptive Statistics of Analysis Variables

Variable	Mean	SD
Treatment Measures		
Center Indicator	0.174	0.379
Primary Care Staff Hours per Week	0.876	4.970
Primary Care Staff Hours per Week (in districts with any center)	4.881	10.761
Primary Care Hours with Hormones Prescribed On Site	2.054	7.442
Primary Care Hours with Hormones Referred, No Condoms	2.019	7.059
Primary Care Hours with Hormones Referred & Condoms Dispensed	0.136	1.389
Primary Care Hours with No Birth Control Services	0.665	3.289
Medical Staff Hours per Week	1.910	10.363
Medical Staff Hours per Week (in districts with any center)	10.987	22.762
Medical Staff Hours with Hormones Prescribed On Site	4.790	16.241
Medical Staff Hours with Hormones Referred, No Condoms	4.149	14.504
Medical Staff Hours with Hormones Referred & Condoms Dispensed	0.466	3.956
Medical Staff Hours with No Birth Control Services	1.569	7.170
Outcome Measures		
Birth Rate per 1,000 Women Aged 15-18	44.28	21.60
10^{th} Grade Dropout Rate (%)	22.39	12.18
11^{th} Grade Dropout Rate (%)	15.43	9.79
12^{th} Grade Dropout Rate (%)	9.28	8.75
14-17 Dropout Rate (%)	10.08	20.59
Female 14-17 Dropout Rate (%)	9.98	20.73
Male 14-17 Dropout Rate (%)	10.16	20.66
18-19 Dropout Rate (%)	15.50	8.20
Female 18-19 Dropout Rate (%)	14.81	7.34
Male 18-19 Dropout Rate (%)	15.82	7.34

Sources: School-based health center service data come from the 1998-2011 National Alliance on School-based Health Care census data. Birth rates are calculated from US vital statistics data from 1990-2012. The 10^{th} through 12^{th} grade dropout rates are calculated from National Center for Education Statistics Common Core of Data on school enrollments and high school diplomas awarded from 1998-2011. The male and female dropout rates come from the 1990 and 2000 US Census as well as the 2005-2011 American Community Survey. Means of treatment variables use the diploma data sample. All service hours are per 1,000 high school aged student in the school district. The "in districts with any center" tabulations showing mean service hours per 1,000 high school aged students among schools districts with any center. Birth control service level means include only those schools districts with any center. All tabulations are school district level means, except for the birth variables which are county level means.

Table 3.3: The Effect of SBHC Services on Teen Birth Rates per 1000 Women

	OLS (i)	2SLS (ii)
Treatment Measure		
Center Indicator	-1.333**	
	(0.474)	
% Effect of Average Center	-3.0%	
Primary Care Staff Hours	-0.218*	-0.492**
	(0.115)	(0.176)
% Effect of Average Center	-2.4%	-5.4%
Medical Staff Hours	-0.108**	-0.228**
	(0.046)	(0.079)
% Effect of Average Center	-2.7%	-5.7%
First-stage F-Stat (Primary Care)		51.34
First-stage F-Stat (Medical Staff)		53.93

Notes: Authors' estimates of equation (1) as described in the text. The dependent variable is 15-18 year old birth rates per 1000. Each cell comes from a separate regression. In column (ii), Primary Care Staff Hours and Medical Staff Hours are instrumented with an indicator for whether there is a center in the county as well as a quadratic in the number of years since a center was first opened in the county (set equal to zero in the years prior to a center first opening). All estimates include county and state-by-year fixed effects, and the regressions are weighted by the high school aged population in the county. Percent effects for the Center Indicator results are calculated by dividing the coefficient by the mean birth rate. The percent effects for the staff hours estimates show the percent effect relative to the mean for a center with the average number of primary care or medical staff hours. Standard errors clustered at the county level are in parentheses: ** indicates significance at the 5% level and * indicates significance at the 10% level.

Table 3.4: The Effect of SBHC Services on Teen Birth Rates per 1000 Women, by Age

	Mother's Age					
	≤14	15	16	17	18	19
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
<u>OLS</u>						
Center Indicator	-0.082**	-0.842**	-1.229**	-0.938*	-1.836**	-1.101**
	(0.025)	(0.258)	(0.409)	(0.504)	(0.595)	(0.560)
% Effect of Average Center	-8.9%	-7.0%	-4.8%	-2.2%	-2.8%	-1.3%
<u>OLS</u>						
Primary Care Staff Hours	-0.016**	-0.129**	-0.201**	-0.222*	-0.257**	-0.194
	(0.007)	(0.052)	(0.119)	(0.113)	(0.131)	(0.128)
% Effect of Average Center	-8.6%	-5.3%	-3.8%	-2.5%	-1.9%	-1.1%
<u>2SLS</u>						
Primary Care Staff Hours	-0.029**	-0.296**	-0.506**	-0.285	-0.684**	-0.374
	(0.008)	(0.091)	(0.142)	(0.184)	(0.259)	(0.257)
% Effect of Average Center	-15.2%	-12.0%	-9.6%	-3.2%	-5.2%	-2.1%
<u>OLS</u>						
Medical Staff Hours	-0.008**	-0.055**	-0.097**	-0.106**	-0.128**	-0.070
	(0.003)	(0.022)	(0.041)	(0.049)	(0.058)	(0.074)
% Effect of Average Center	-9.1%	-5.0%	-4.1%	-2.7%	-2.2%	-0.9%
<u>2SLS</u>						
Medical Staff Hours	-0.013**	-0.132**	-0.230**	-0.136*	-0.319**	-0.186
	(0.004)	(0.041)	(0.064)	(0.083)	(0.116)	(0.115)
% Effect of Average Center	-15.0%	-12.1%	-9.8%	-3.5%	-5.4%	-2.4%
Mean Birth Rate	0.93	11.98	25.85	43.41	64.57	86.31

Notes: Authors' estimates of equation (1) as described in the text. Each cell comes from a separate regression. The *Center Indicator* IV estimates instrument Primary Care Staff Hours or Medical Staff Hours with an indicator for whether there is a center in the county as well as a quadratic in the number of years since a center was first opened in the county (set equal to zero in the years prior to a center first opening). All estimates include county and state-by-year fixed effects, and the regressions are weighted by the high school aged population in the county. Percent effects for the Center Indicator results are calculated by dividing the coefficient by the mean birth rate for that age group. The percent effects for the staff hours estimates show the percent effect relative to the mean for a center with the average number of primary care or medical staff hours. Standard errors clustered at the county level are in parentheses: ** indicates significance at the 5% level and * indicates significance at the 10% level.

Table 3.5: The Effect of SBHC Services on Teen Birth Rates per 1000 Women, by Birth Control Services

Service Measure: Birth Control Services	Primary Care Staff Hours (i)	Medical Staff Hours (ii)
Hormones Prescribed On Site	-0.628**	-0.239*
	(0.271)	(0.134)
Hormones Referred, No Condoms	-0.244	-0.144
	(0.161)	(0.093)
Hormones Referred &	-0.221	-0.108
Condoms Dispensed	(0.551)	(0.113)
No Birth Control Services	-0.567	-0.238**
	(0.363)	(0.079)

Notes: Authors' estimates of equation (1) as described in the text. The dependent variable is 15-18 year old birth rates per 1000. Each column comes from a separate regression. The birth control service measures include the number of service hours of each type in centers with the given birth control policy. The birth control policy groups are exhaustive and mutually exclusive. All estimates include county and state-by-year fixed effects, and the regressions are weighted by the high school aged population in the county. Standard errors clustered at the county level are in parentheses: ** indicates significance at the 5% level and * indicates significance at the 10% level.

Table 3.6: The Effect of SBHC Services on Teen Birth Rates per 1000 Women, by Race/Ethnicity

	OLS			2SLS		
Race/Ethnicity:	White (i)	Black (ii)	Hispanic (iii)	White (iv)	Black (v)	Hispanic (vi)
Treatment Measure						
Center Indicator	0.138	-1.400*	-3.204*			
	(0.244)	(0.828)	(1.757)			
% Effect of Average Center	0.7%	-2.4%	-4.7%			
Primary Care Staff Hours	0.006	-0.157	-0.677*	-0.026	-0.928**	-1.113
	(0.052)	(0.157)	(0.373)	(0.096)	(0.375)	(0.773)
% Effect of Average Center	0.2%	-1.3%	-4.8%	-0.6%	-7.9%	-7.9%
Medical Staff Hours	0.004	-0.120*	-0.410**	-0.011	-0.408**	-0.532
	(0.019)	(0.067)	(0.170)	(0.043)	(0.168)	(0.347)
% Effect of Average Center	0.2%	-2.3%	-6.6%	-0.6%	-7.8%	-8.5%
Mean Birth Rate	19.60	57.66	68.35	19.60	57.66	68.35

Notes: Authors' estimates of equation (1). The dependent variable is 15-18 year old birth rates per 1000. Each cell comes from a separate regression. In columns (iv)-(vi), Primary Care Staff Hours and Medical Staff Hours are instrumented with an indicator for whether there is a center in the county as well as a quadratic in the number of years since a center was first opened in the county (set equal to zero in the years prior to a center first opening). All estimates include county and state-by-year fixed effects, and the regressions are weighted by the high school aged population in the county. Percent effects for the Center Indicator results are calculated by dividing the coefficient by the mean birth rate for the age and racial/ethnic group. The percent effects for the staff hours estimates show the percent effect relative to the mean for a center with the average number of primary care or medical staff hours. Standard errors clustered at the county level are in parentheses: ** indicates significance at the 5% level and * indicates significance at the 10% level.

Table 3.7: The Effect of SBHC Services on Teen Birth Rates – Robustness Checks

	State-Year-Median Income Fixed Effects		Year-Bottom 20% Income Fixed Effects		Allowing for Level Shifts
	OLS	IV	OLS	IV	OLS
Treatment Measure	(i)	(ii)	(iii)	(iv)	(v)
Center Indicator	-1.238**		-1.280**		
	(0.478)		(0.492)		
% Effect of Average Center	-2.8%		-2.9%		
Primary Care Staff Hours	-0.216**	-0.648**	-0.193*	-0.506**	-0.121
	(0.107)	(0.174)	(0.105)	(0.173)	(0.118)
Center Indicator					-1.120**
					(0.535)
% Effect of Average Center	-2.4%	-7.1%	-2.2%	-5.6%	-4.0%
Medical Staff Hours	-0.110**	-0.288**	-0.097**	-0.234**	-0.060
	(0.042)	(0.077)	(0.040)	(0.078)	(0.052)
Center Indicator					-1.089**
					(0.529)
% Effect of Average Center	-2.7%	-7.1%	-2.5%	-5.8%	-4.0%
First-stage F-Stat (Primary Care)		50.45		52.09	
First-stage F-Stat (Medical Staff)		52.64		54.81	

Notes: Authors' estimation as described in the text. The dependent variable is 15-18 year old birth rates per 1000. *Center Indicator* is an indicator variable equal to 1 if any school-based health center exists in the county. All results contain county fixed effects. Estimates in columns (i) and (ii) include state-year-median income fixed effects, where median income is an indicator for whether the 1990 median household income in the county is above the median household income in the state in 1990. Estimates in columns (iii) and (iv) include state-year fixed effects and year fixed effects interacted with an indicator for the county being in the bottom 20% of median household income in 1990. Estimates in column (v) contain state-year fixed effects. Regressions are weighted by the high school aged population in the county. Standard errors clustered at the county level are in parentheses: ** indicates significance at the 5% level and * indicates significance at the 10% level.

Table 3.8: The Effect of SBHC Services on Birth Rates Among Older Women Without Access to a SBHC

	Mother's Age		
	20-24	25-29	15-18
Treatment Measure	(i)	(ii)	(iii)
Center Indicator	0.421	-0.067	-1.674*
	(1.130)	(0.770)	(0.906)
Primary Care Staff Hours	0.101	0.399**	-0.454**
	(0.128)	(0.115)	(0.178)
Medical Staff Hours	0.023	0.138**	-0.123*
	(0.059)	(0.059)	(0.075)
Mean Birth Rate	100.58	114.19	44.28

Notes: Authors' estimates of equation (1) as described in the text. Each cell comes from a separate regression. The sample consists of states in which less than half of centers report they are accessible to those who do not attend the school in which they are located and are restricted to counties in which there were no centers when the women in each age group were of high school age. All estimates include county and state-by-year fixed effects, and the regressions are weighted by the high school aged population in the county. Standard errors clustered at the county level are in parentheses: ** indicates significance at the 5% level and * indicates significance at the 10% level.

Table 3.9: The Effect of SBHC Services on High School Dropout Rates (in Percent) – Diploma Data

	OLS			2SLS		
Grade:	10^{th}	11^{th}	12^{th}	10^{th}	11^{th}	12^{th}
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
<u>Treatment Measure</u>						
Center Indicator	0.576	0.014	0.600			
	(0.426)	(0.529)	(0.442)			
Primary Care Staff Hours	-0.006	-0.023	-0.003	0.064	-0.024	0.154*
	(0.012)	(0.020)	(0.013)	(0.086)	(0.097)	(0.091)
% Effect of Average Center	-0.03%	-0.13%	-0.01%	0.31%	-0.12%	0.75%
Medical Staff Hours	-0.005	-0.010	-0.003	0.042	-0.003	0.068
	(0.006)	(0.007)	(0.006)	(0.042)	(0.048)	(0.042)
% Effect of Average Center	-0.02%	-0.05%	-0.01%	0.20%	-0.02%	0.33%
First-stage F-Stat (Primary Care)					44.18	
First-stage F-Stat (Medical Staff)					45.51	

Notes: Authors' estimates of equation (4) using NCES CCD high school diploma data from 1998-2010. Each cell comes from a separate regression. In columns (iv)-(vi), Primary Care Staff Hours and Medical Staff Hours are instrumented with an indicator for whether there is a center in the school district as well as a quadratic in the number of years since a center was first opened in the school district (set equal to zero in the years prior to a center first opening). The 10^{th} grade dropout rate is calculated as 1 minus the ratio of diplomas awarded in year t and the 10^{th} grade enrollment in year $t-2$. The 11^{th} grade dropout rate equals 1 minus the ratio of diplomas awarded in year t and the 11^{th} grade enrollment in year $t-1$, and the 12^{th} grade dropout rate is calculated as 1 minus the ratio of diplomas awarded in year t and the 12^{th} grade enrollment in year t. All estimates include school district and state-by-year fixed effects, and the regressions are weighted by the high school aged population in the school district. The percent effects for the staff hours estimates show the percent effect for a center with the average number of primary care or medical staff hours. Standard errors clustered at the school district level are in parentheses: ** indicates significance at the 5% level and * indicates significance at the 10% level.

Table 3.10: The Effect of SBHC Services on High School Dropout Rates (in Percent) – Census/ACS Data

	14-17 Year Olds			18-19 Year Olds		
	All	Female	Male	All	Female	Male
Treatment Measure	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Center Indicator	-0.124	-0.123	-0.141	0.484	0.234	0.682
	(0.148)	(0.169)	(0.155)	(0.503)	(0.610)	(0.555)
Primary Care Staff Hours	-0.005	-0.002	-0.007	-0.011	-0.052*	0.013
	(0.005)	(0.006)	(0.005)	(0.020)	(0.028)	(0.024)
% Effect of Average Center	-0.02%	-0.01%	-0.03%	-0.05%	-0.25%	0.06%
Medical Staff Hours	-0.002	-0.001	-0.003	-0.002	-0.025*	0.018
	(0.002)	(0.003)	(0.003)	(0.010)	(0.015)	(0.012)
% Effect of Average Center	-0.01%	-0.01%	-0.01%	-0.01%	-0.12%	0.04%

Notes: Authors' estimates of equation (4) using 1990 and 2000 Census data as well as 2005-2011 ACS data. Each cell comes from a separate regression. The dropout rates measure the proportion of each age group living in the district that does not report attending school and that does not have a high school degree. All estimates include school district and state-by-year fixed effects, and the regressions are weighted by the high school aged population in the school district. The percent effects for the staff hours estimates show the percent effect for a center with the average number of primary care or medical staff hours. Standard errors clustered at the school district level are in parentheses: ** indicates significance at the 5% level and * indicates significance at the 10% level.

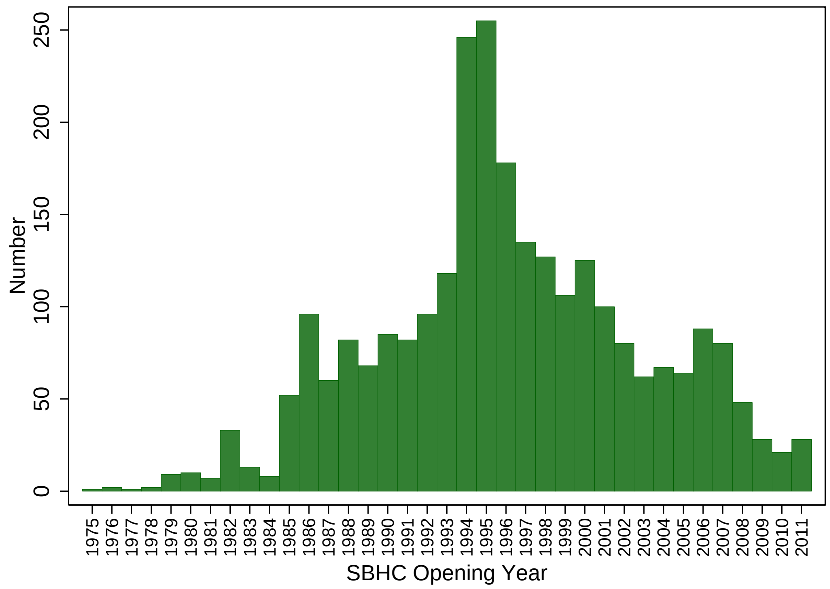
Table 3.11: The Relationship Between SBHC Services and Per-Student Expenditures

Treatment Measure	Log Per Student Expenditures
Center Indicator	-0.003
	(0.006)
Primary Care Staff Hours	0.0005
	(0.0003)
Medical Staff Hours	0.0002
	(0.0001)

Notes: Authors' estimation as described in the text using data from the 1998-2011 Common Core of Data. All estimates include school district and state-by-year fixed effects. Regressions are weighted by the high school aged population in the school district. Standard errors clustered at the school district level are in parentheses: ** indicates significance at the 5% level and * indicates significance at the 10% level.

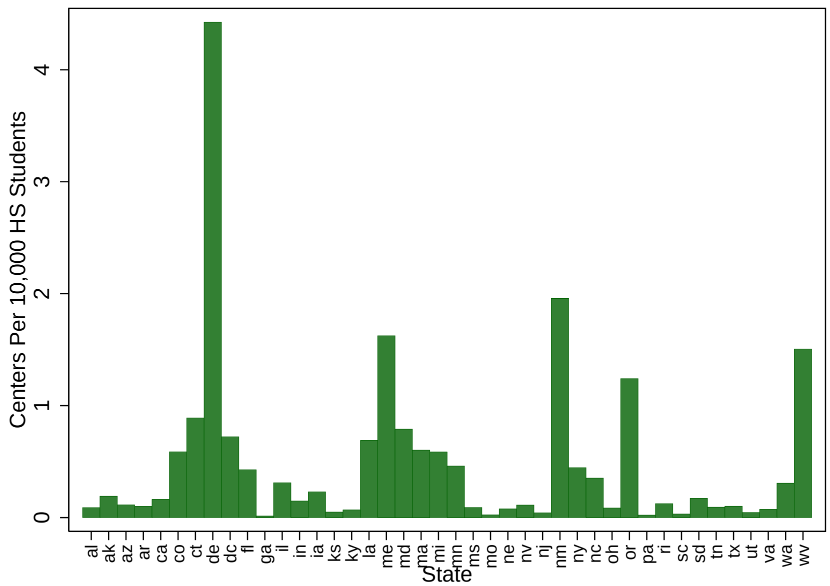
3.8 Figures

Figure 3.1: Distribution of SBHC Opening Years



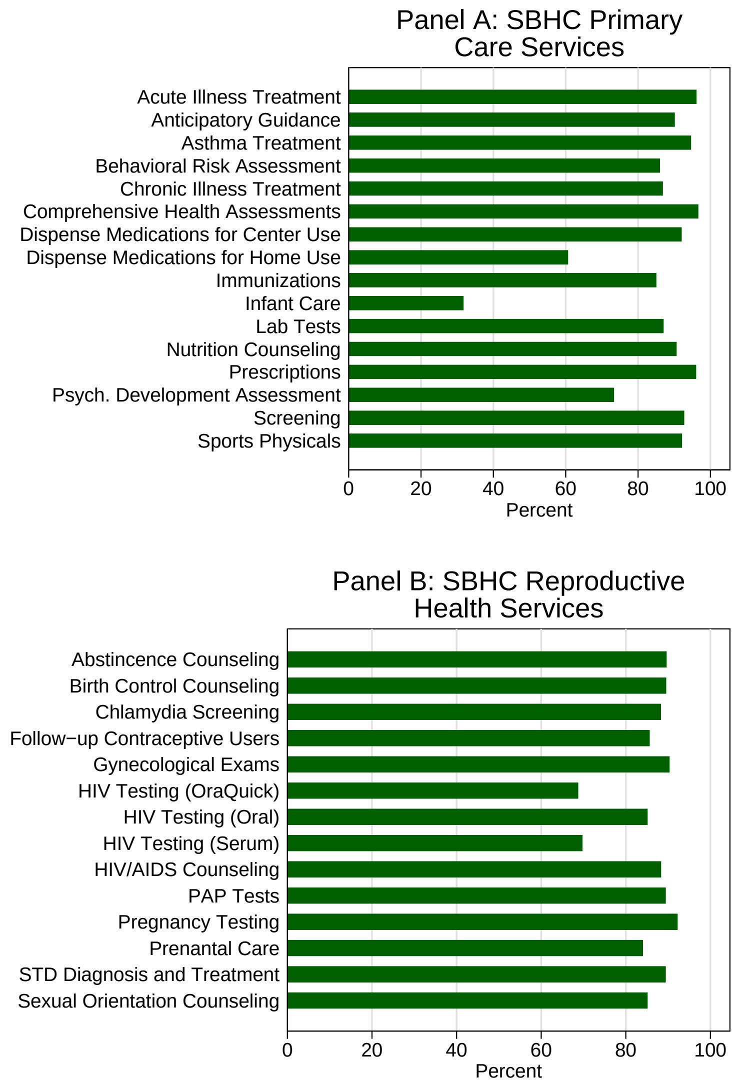
Source: NASBHC School-based Health Center Census, 1998-2011. The figure includes only centers that serve high school students.

Figure 3.2: Distribution of SBHCs per 10,000 High School Students Across States



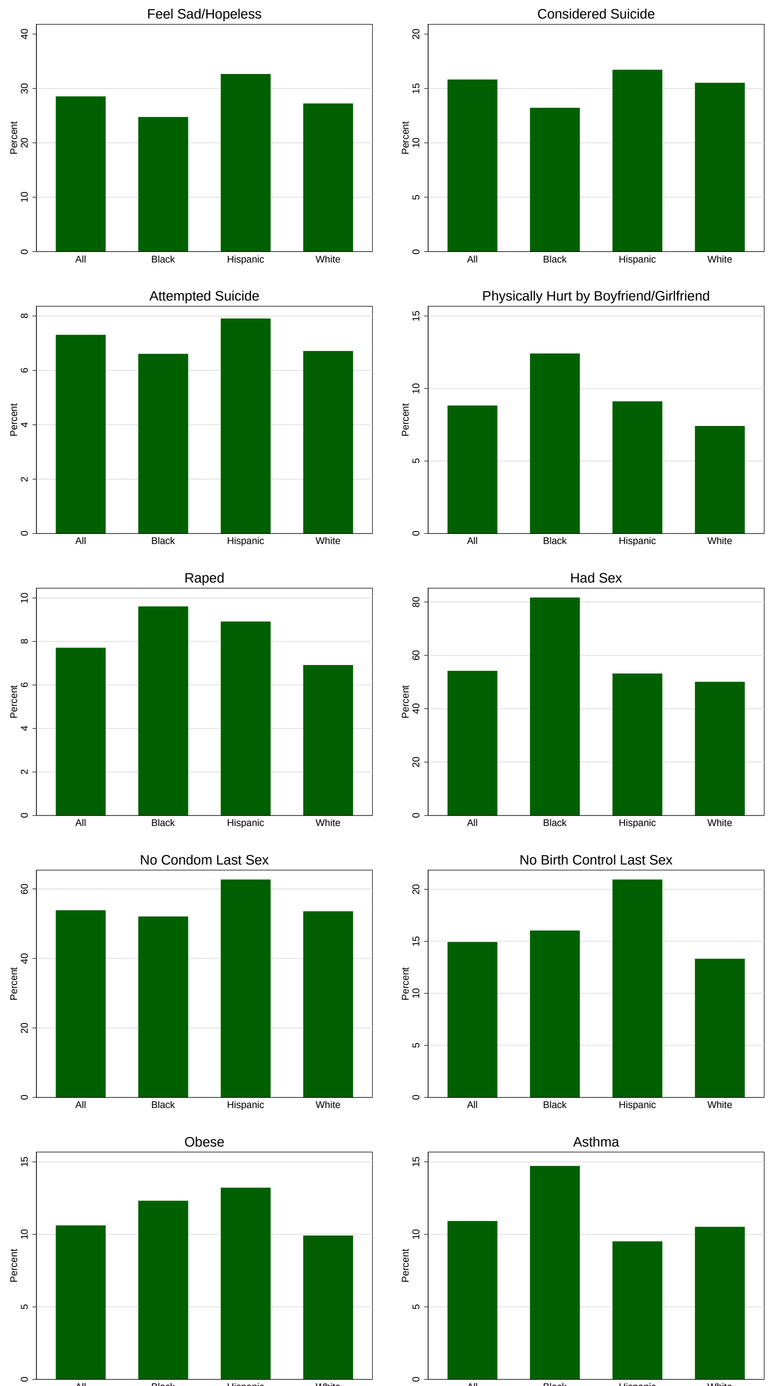
Source: NASBHC School-based Health Center Census, 2011. The figure includes only centers that serve high school students. States without SBHCs (HI, ID, MT, ND, NH, SD, VT, WI, WY) are omitted from the figure.

Figure 3.3: Primary Care and Non-Contraceptive Reproductive Health Services Provided by SBHCs



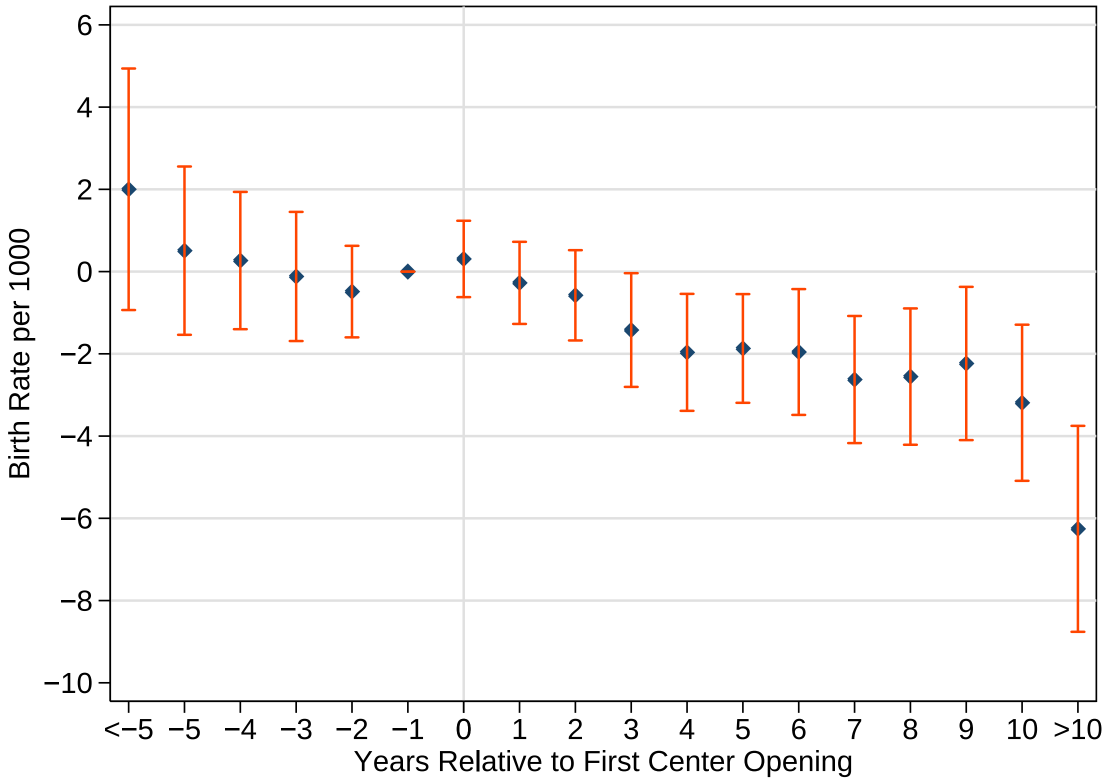
Source: These figures are reproduced from the 2007-2008 School-based Health Centers National Census annual report, available at http://www.sbh4all.org/atf/cf/%7Bcd9949f2-2761-42fb-bc7a-cee165c701d9%7D/NASBHC%202007-08%20CENSUS%20REPORT%20FINAL.PDF. The reproductive care service tabulations show the percent providing each service on-site and the percent providing referrals for each service.

Figure 3.4: Health Outcomes Among High-School-Aged Students, 2011 YRBSS



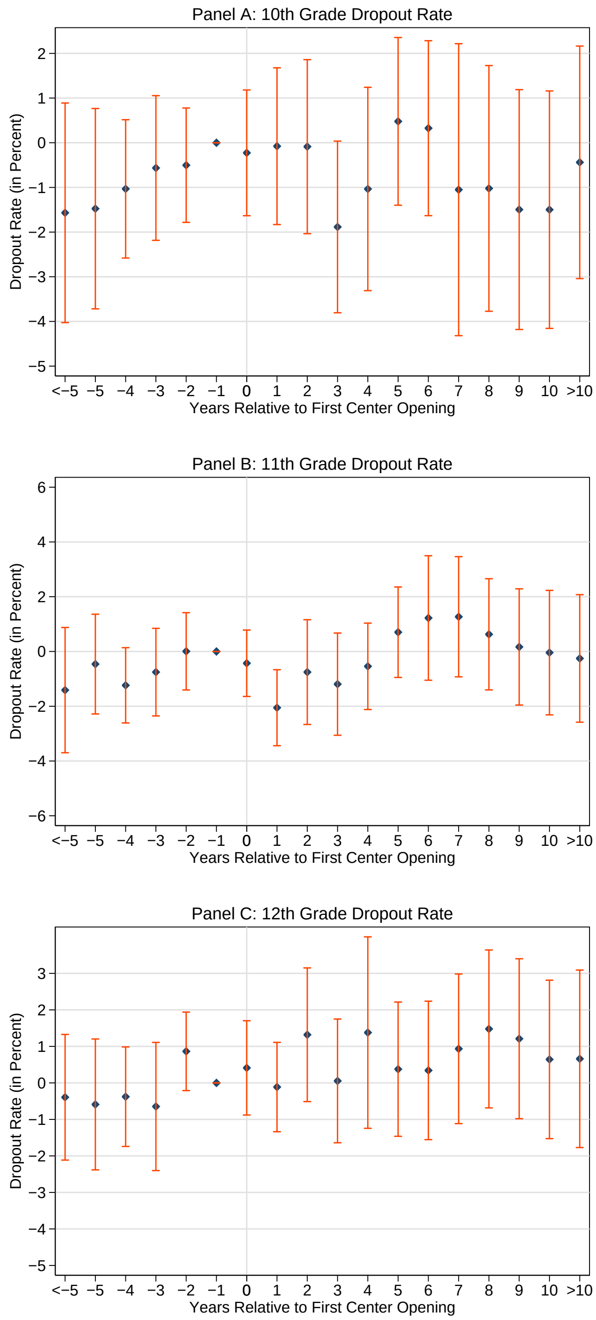
Source: 2011 Youth Risk Behavior Surveillance System (YRBSS).

Figure 3.5: Event Study Estimates of the Effect of SBHC Entry on Teen Birth Rates (per 1000 women)



Authors' estimates of equation (2) as described in the text. The dependent variable is 15-18 year old birth rates per 1000. Each point shows the coefficient estimate on the service measure or center indicator interacted with the relative time to the first center opening in the county. All estimates include county and state-by-year fixed effects, and the regressions are weighted by the high school aged population in the county. The lines extending from each point show the bounds of the 95% confidence intervals that are calculated using standard errors clustered at the county level. Relative year -1 is omitted, so all estimates are relative to this year.

Figure 3.6: Event Study Estimates of the Effect of SBHC Entry on High School Dropout Rates (in Percent) – Diploma Data



Authors' estimates of equation (5) as described in the text. Each point shows the coefficient estimate on the total medical staff hours service measure interacted with the relative time to the first center opening in the school district. All estimates include school district and state-by-year fixed effects, and the regressions are weighted by the high school aged population in the school district. The lines extending from each point show the bounds of the 95% confidence intervals that are calculated using standard errors clustered at the school district level. Relative year -1 is omitted, so all estimates are relative to this year.

BIBLIOGRAPHY

[1] Acemoglu, Daron, and Joshua Angrist. 2000. "How Large Are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws." *National Bureau of Economic Research Macroeconomics Annual 2000*, edited by Ben S. Bernanke and Kenneth Rogoff, pp. 9 59. MIT Press.

[2] Adler, Nancy E. and David H. Rehkopf. 2008. "U.S. Disparities in Health: Descriptions, Causes, and Mechanisms." *Annual Review of Public Health* 29: 235–252.

[3] Advocates for Children of New Jersey. 2016. "Showing Up Matters: The State of Chronic Absenteeism in New Jersey, 2nd Annual Report."

[4] Akerlof, George A. and Rachel E. Kranton. 2002. Identity and Schooling: Some Lessons for the Economics of Education. *Journal of Economic Literature*, 40(4): 1167-1201.

[5] Alexander, K. L., Entwisle, D. R., & Olson, L. S. 2007. "Lasting consequences of the summer learning gap." *American Sociological Review,* 72(2), 167-180.

[6] Almond, Douglas and Janet Currie. 2011. "Killing Me Softly: The Fetal Origins Hypothesis." *Journal of Economic Perspectives* 25(3): 153-172.

[7] Ananat, Elizabeth Oltmans and Daniel M. Hungerman. 2012. "The Power of the Pill for the Next Generation: Oral Contraception's Effects on Fertility, Abortion, and Maternal and Child Characteristics." *Review of Economics and Statistics* 94(1): 37–51.

[8] Andrulis, Dennis P. 1998. "Access to Care Is the Centerpiece in the Elimination of Socioeconomic Disparities in Health." *Annals of Internal Medicine* 129(5): 412-416.

[9] Angrist, J. D., & Keueger, A. B. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?." *The Quarterly Journal of Economics, 106*(4), 979–1014.

[10] Archambault, I., Janosz, M., Fallu, J. S., & Pagani, L. S. 2009. "Student engagement and its relationship with early high school dropout." *Journal of adolescence,* 32(3), 651-670.

[11] Arcidiacono, Peter, Ahmed Khwaja, and Lijing Ouyang. 2012. "Habit Persistence and Teen Sex: Could Increased Access to Contraception Have Unintended Consequences for Teen Pregnancies?." *Journal of Business and Economic Statistics* 30(2): 312-325.

[12] Arulampalam, W., Gregg, P., & Gregory, M. 2001. "Unemployment scarring." *The Economic Journal*, 111(475), 577-584.

[13] Aschraft, Adam, Ivan Fernandez-Val and Kevin Lang. 2013. "The Consequences of Teenage Childbearing: Consistent Estimates When Abortion Makes Miscarriage Non-random." *The Economic Journal* 123(571): 875-905.

[14] Ashenfelter, Orley, and Alan Krueger. 1994. "Estimating the Returns to Schooling Using a New Sample of Twins." *American Economic Review* 84 (December): 1157 73.

[15] Ashenfelter, Orley, and Cecilia Elena Rouse. 1998. "Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins." *Quarterly Journal of Economics* 113 (February): 253 84.

[16] Astone, N. M., & McLanahan, S. S. 1991. "Family structure, parental practices and high school completion." *American sociological review,* 309-320.

[17] Attendance Works. 2015." NEW FEDERAL EDUCATION LAW INCLUDES CHRONIC ABSENCE TRACKING, TRAINING," web: ¡http://www.attendanceworks.org/new-federal-education-law-includes-chronic-absence-tracking-training/¿

[18] Bailey, Martha J. 2006. "More Power to the Pill: The Impact of Contraceptive Freedom on Women's Life Cycle Labor Supply." *Quarterly Journal of Economics* 121(1): 289–320.

[19] Bailey, Martha J. 2010. ""Momma's Got the Pill": How Anthony Comstock and Griswold v. Connecticut Shaped US Childbearing." *American Economic Review* 100(1): 98-129.

[20] Bailey, Martha J. and Andrew Goodman-Bacon. 2015. "The War on Poverty's Experiment in Public Medicine: The Impact of Community Health Centers on the Mortality of Older Americans." *American Economic Review* 105(3): 1067-1104.

[21] Balfanz, R.; Byrns, V. 2014. "Meeting the Challenge of Combating

Chronic Absenteeism." *Everyone Graduates Center Report, Johns Hopkins University.* web: ¡http://www.attendanceworks.org/wordpress/wp-content/uploads/2014/01/NYC-Chronic-Absenteeism-Impact-Report-Nov-2013.pdf¿

[22] Balfanz, R.; Herzog, L.; Mac Iver, D.J. 2007. "Preventing Student Disengagement and Keeping Students on the Graduation Path in Urban Middle-Grades Schools: Early Identification and Effective Interventions." *Educational Psychologist*, 42(4) 223-235.

bibitem Bleakley, Hoyt. 2007. "Disease and Development: Evidence from Hookworm Eradication in the American South." *The Quarterly Journal of Economics* 122(1): 73-117.

[23] Bleakley, Hoyt. 2010. "Malaria Eradication in the Americas: A retrospective Analysis of Childhood Exposure." *American Economic Journal: Applied Economics* 2(2): 1-45.

[24] Becker, Gary S. 1967. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Columbia University Press.

[25] Belfield, C. R., & Levin, H. M. (Eds.). 2008. *The Price We Pay : Economic and Social Consequences of Inadequate Education*. New York, NY, US: Brookings Institution Press.

[26] Blakemore, S.J. & Robbins, T.W., 2012. "Decision-making in the adolescent brain." *Nature neuroscience*, 15(9), pp.1184-1191.

[27] Bridgeland, J. M., DiIulio Jr, J. J., & Morison, K. B. 2006. "The silent epidemic: Perspectives of high school dropouts." *Civic Enterprises.*

[28] Brown, David, Amanda E. Kowalski, and Ithai Lurie. 2014. "Medicaid as an Investment in Children: What is the Long-Term Impact on Tax Receipts?" Yale University Working Paper, available at http://www.econ.yale.edu/ ak669/medicaid.latest.draft.pdf, last accessed 10/7/2014.

[29] Buckles, Kasey and Daniel Hungerman. 2014. "Estimating the Incidental Fertility Effects of School Condom Distribution Programs." Mimeo.

[30] Cameron, S. V., & Heckman, J. J. 1993. "The Nonequivalence of High School Equivalents." *Journal of Labor Economics*, 1-47.

[31] Cameron, S.V. and Heckman, J.J., 1998. "Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males." The Journal of Political Economy, 106(2), p.262.

[32] Card, D. 2001. "Estimating the return to schooling: Progress on some persistent econometric problems." *Econometrica, 69*(5), 1127-1160.

[33] Case, Ann, Darren Lubotsky, and Christina Paxson. 2002. "Economic Status and Health in Childhood: The Origins of the Gradient." *American Economic Review* 92(5): 1308–1334.

[34] Case, Anne, Angela Fertig, and Christina Paxson. 2005. "The Lasting Impact of Childhood Health and Circumstance." *Journal of Health Economics* 24(2): 365-389.

[35] Casey, B.J., Jones, R.M. and Hare, T.A., 2008. "The adolescent brain." *Annals of the New York Academy of Sciences*, 1124(1), pp.111-126.

[36] Coates, D. 2003. "Education production functions using instructional time as an input." *Education Economics*, 11(3), 273-292.

[37] Cohodes, Sarah, Daniel Grossman, Samuel Kleiner, and Michael F. Lovenheim. Forthcoming. "The Effect of Child Health Insurance Access on Schooling: Evidence from Public Insurance Expansions." *Journal of Human Resources*.

[38] Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. 1996. "The effects of summer vacation on achievement test scores: A narrative and meta-analytic review." *Review of educational research*, 66(3), 227-268.

bibitem Conti, Gabriella, James Heckman and Sergio Urzua. 2010. "The Education-Health Gradient." *American Economic Review* 100(2): 234–238.

[39] Cuhna, Flavio, James J. Heckman, Lance Lochner, and Dimitriy Masterov. 2006. "Interpreting the Evidence on Life Cycle Skill Formation." In *Handbook of the Economics of Education*, Vol. 1, ed. Eric Hanushek and Finis Welch, 697-812. Amsterdam: Elsevier.

[40] Currie, Janet and Jonathan Gruber. 1996. "Health Insurance Eligibility, Utilization of Medical Care, and Child Health." *Quarterly Journal of Economics* 111(2): 431–466.

[41] Currie, Janet, Sandra Decker and Wanchuan Lin. 2008. "Has Public Health

Insurance For Older Children Reduced Disparities in Access to Care and Health Outcomes?" *Journal of Health Economics* 27(6): 1567–1581.

[42] Dahl, Gordon B. and Lance Lochner. 2012. "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit." *American Economic Review* 102(5): 1927–1956.

[43] Dobkin, C. and Ferreira, F., 2010. "Do school entry laws affect educational attainment and labor market outcomes?" *Economics of Education Review,* 29(1), pp.40-54

[44] Downey, D. B., Von Hippel, P. T., & Broh, B. A. (2004). "Are schools the great equalizer? Cognitive inequality during the summer months and the school year." *American Sociological Review,* 69(5), 613-635.

[45] Allensworth, E.M. & Easton, J.Q. 2007. "What Matters for Staying On-Track and Graduating in Chicago Public High Schools." *Consortium on Chicago School Research, University of Chicago: Research Report*

[46] Edin, P. A., & Gustavsson, M. 2008. "Time out of work and skill depreciation." *ILR Review, 61(2)*, 163-180.

[47] Epstein, J. L., & Sheldon, S. B. 2002. "Present and accounted for: Improving student attendance through family and community involvement." *The Journal of Educational Research*, 95(5), 308-318.

[48] Evans, W.N., Morrill, M.S. and Parente, S.T., 2010. "Measuring inappropriate medical diagnosis and treatment in survey data: The case of ADHD among school-age children." Journal of health economics, 29(5), pp.657-673.

[49] Figlio, David, Guryan, Jonathan, Karbownik, Krzysztof, and Jeffrey Roth. 2014. "The Effects of Poor Neonatal Health on Children's Cognitive Development." *American Economic Review* 104(12): 3921-55.

[50] Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker and Oregon Health Study Group. 2012. "The Oregon Health Insurance Experiment: Evidence from the First Year." *Quarterly Journal of Economics* 127(3): 1057–1106.

[51] Fletcher, Jason M. and Barbara L. Wolfe. 2009. "Education and Labor Market Consequences of Teenage Childbearing: Evidence Using the Timing of

Pregnancy Outcomes and Community Fixed Effects." *Journal of Human Resources* 44(2): 303–325.

[52] Flores, Glenn. 2010. "Technical Report: Racial and Ethnic Disparities in the Health and Health Care of Children." *Pediatrics* 125(4): e979–e1020.

[53] Gross, Tal and Matthew J. Notowidigdo. 2011. "Health Insurance and the Consumer Bankruptcy Decision: Evidence from Expansions of Medicaid." *Journal of Public Economics* 95(7-8): 767–778.

[54] Geierstanger, Sara Peterson, Gorette Amaral, Mona Mansour, and Susan Russell Walters. 2004. "School-Based Health Centers and Academic Performance: Research, Challenges, and Recommendations." *Journal of School Health* 74(9): 347-352.

[55] Geronimus, Arlene T. and Sanders Korenman. 1992. "The Socioeconomic Consequences of Teen Childbearing Reconsidered." *Quarterly Journal of Economics* 107(4): 1187–1214.

[56] Goldin, Claudia and Lawrence F. Katz. 2002. "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions." *Journal of Political Economy* 110(4): 730–770.

[57] Hanushek, E. A. (2002). Publicly provided education. *Handbook of public economics*, 4, 2045-2141.

[58] Harris, Kathleen Mullan, Penny Gordon-Larsen, Kim Chantala, and J. Richard Udry. 2006. "Longitudinal Trends in Race/Ethnic Disparities in Leading Health Indicators From Adolescence to Young Adulthood." *Archives of Pediatric Adolescent Medicine* 160: 74–81.

[59] Heckman, James J. and Paul A. LaFontaine. 2010. "The American High School Graduation Rate: Trends and Levels." *Review of Economics and Statistics* 92(2): 244–262.

[60] Heckman, James J. and Paul A. LaFontaine. 2006. "Bias-corrected Estimates of GED Returns." *Journal of Labor Economics* 24(3): 661–700.

[61] Heckman, J.J., Stixrud, J. and Urzua, S., 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics*, 24(3), pp.411-482.

[62] Heckman, J.J., 2006. "Skill formation and the economics of investing in disadvantaged children." *Science*, 312(5782), pp.1900-1902.

[63] Heckman, J. J. 2000. "Policies to foster human capital." *Research in economics,* 54(1), 3-56.

[64] Hotz, Joseph V., Susan William McElroy and Seth G. Sanders. 2005. "Teenage Childbearing and Its Life Cycle Consequences: Exploiting a Natural Experiment." *Journal of Human Resources* 40(3): 683–715.

[65] Herzog, S. 2005. "Measuring determinants of student return vs. dropout/stopout vs. transfer: A first-to-second year analysis of new freshmen." *Research in higher education*, 46(8), 883-928.

[66] Imberman, S. A. 2011. "Achievement and behavior in charter schools: Drawing a more complete picture." *The Review of Economics and Statistics,* 93(2), 416-435.

[67] Jacob, B.A. and Lefgren, L., 2003. "Are idle hands the devil's workshop? Incapacitation, concentration, and juvenile crime." *The American Economic Review*, 93(5), pp.1560-1577.

[68] Kaestner, Robert, T. Joyce and A. Racine. 2001. "Medicaid Eligibility and the Incidence of Ambulatory Care Sensitive Hospitalizations for Children." *Social Science and Medicine* 52(2): 305-313.

[69] Kahneman, Daniel. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics." American Economic Review, 93(5): 1449-1475.

[70] Kearney, C. A. 2008. "School absenteeism and school refusal behavior in youth: A contemporary review." *Clinical psychology review*, 28(3), 451-471.

[71] Kearney, Melissa S. and Phillip B. Levine. 2009. "Subsidized Contraception, Fertility, and Sexual Behavior." *The Review of Economics and Statistics* 91(1): 137-151.

[72] Kearney, Melissa S. and Phillip B. Levine. 2012. "Why is the Teen Birth Rate in the United States So High and Why Does It Matter?" *Journal of Economic Perspectives* 26(2): 141-166.

[73] Kearney, Melissa S. and Phillip B. Levine. 2014. "Media Influences on So-

cial Outcomes: The Impact of MTV's *16 and Pregnant* on Teen Childbearing."NBER Working Paper No. 19795.

[74] Kearns, Suzanne E.U., Michael D. Pullmann, Sarah Cusworth Walker, Aaron R. Lyons, T.J. Cosgrove, and Eric J. Bruns. 2011. "Adolescent Use of School-Based Health Centers and High School Dropout." *Archives of Pediatrics & Adolescent Medicine* 165(7): 617-623.

[75] Kisker, Ellen E. and Randall S. Brown. 1996. "Do School-Based Health Centers Improve Adolescents' Access to Health Care, Health Status, and Risk-Taking Behavior?" *Journal of Adolescent Health* 18(5): 335-343.

[76] Klepinger, Daniel, Shelly Lundberg and Robert Plotnick. 1999. "How Does Adolescent Fertility Affect the Human Capital and Wages of Young Women?" *Journal of Human Resources* 34(3): 421-448.

[77] Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics* 114 (May): 497 531.

[78] Lamdin, D. J. 1996. "Evidence of student attendance as an independent variable in education production functions." *The Journal of educational research,* 89(3), 155-162.

[79] Lazear, E. P. 2001. "Educational production." *The Quarterly Journal of Economics,* 116(3), 777-803.

[80] Leininger, Lindsey, Helen Levy and Diane Schanzenbach. 2009. "Consequences of SCHIP Expansions for Household Well-Being." *Forum for Health Economics & Policy (Frontiers in Health Policy Research)* 13(1), Article 3.

[81] Levine, David I. and Gary Painter. 2003. "The Schooling Costs of Teenage Out-of-Wedlock Childbearing: Analysis with a Within-school Propensity-score-matching Estimator." *The Review of Economics and Statistics* 85(4): 884–900.

[82] Levitt, S. D., List, J. A., Neckermann, S., & Sadoff, S. 2016. "The behavioralist goes to school: Leveraging behavioral economics to improve educational performance." *American Economic Journal: Economic Policy*, 8(4), 183-219.

[83] Link, Bruce G., and Jo C. Phelan. 1995. "Social Conditions as Fundamental Causes of Disease." *Journal of Health and Social Behavior*, special number: 80 94.

[84] Lleras-Muney, A., 2005. “The relationship between education and adult mortality in the United States.” *The Review of Economic Studies*, 72(1), pp.189-221.

[85] Leos-Urbel, J., Schwartz, A.E., Weinstein, M. and Corcoran, S., 2013. “Not just for poor kids: The impact of universal free school breakfast on meal participation and student outcomes.” *Economics of education review*, 36, pp.88-107.

[86] Lochner, Lance, and Enrico Moretti. 2004. “The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports.” *American Economic Review* 94, no. 1: 155 89.

[87] Marburger, D. R. 2006. “Does mandatory attendance improve student performance?.” *The Journal of Economic Education,* 37(2), 148-155.

[88] Martin, Joyce A., Brady E. Hamilton, Stephanie J. Ventura, Michelle JK Osterman, and T. J. Mathews. 2013. “Births: Final Data for 2011.” National Vital Statistics Report 62, no. 1: 1-90.

[89] McCrary, J. and Royer, H., 2011. “The effect of female education on fertility and infant health: Evidence from school entry policies using exact date of birth.” *The American economic review, 101(1), pp.158-195.*

[90] Michelmore, Katherine. 2013. “The Effect of Income on Educational Attainment: Evidence from State Earned Income Tax Credit Expansions.” Mimeo.

[91] Miguel, Edward, and Michael Kremer. 2004. “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities.” *Econometrica* 72(1): 159-217.

[92] Mincer, Jacob. 1974. *Schooling, Experience, and Earnings*. Columbia University Press.

[93] Mishel, Lawrence and Joydeep Roy. 2006. *Rethinking High School Graduation Rates and Trends*. Washington, D.C.: Economic Policy Institute.

[94] Murnane, Richard J. 2013. “U.S. High School Graduation Rates: Patterns and Explanations.” *Journal of Economic Literature* 51(2): 370–422.

[95] Neild, R. C., Stoner-Eby, S., & Furstenberg, F. 2008. “Connecting entrance and departure: The transition to ninth grade and high school dropout.” *Education and Urban Society*, 40(5), 543-569.

[96] Olson, L.S. 2014. "Why September Matters: Improving Student Attendance." *Baltimore Education Research Consortium Policy Brief.* web: ¡http://baltimore-berc.org/wp-content/uploads/2014/08/SeptemberAttendanceBriefJuly2014.pdf.

[97] Oreopoulos, P. 2007. "Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling." *Journal of public Economics, 91*(11), 2213-2229.

[98] Lavecchia, A. M., Liu, H., & Oreopoulos, P. 2014. "Behavioral economics of education: Progress and possibilities" (No. w20609). National Bureau of Economic Research.

[99] Reback, Randall and Tamara Lalovic Cox. 2016. "Where Health Policy Meets Education Policy: School-based Health Centers in New York City." Mimeo.

[100] Reid, K. 2005. "The causes, views and traits of school absenteeism and truancy: An analytical review." *Research in Education*, 74(1), 59-82.

[101] Ribar, David C. 1994. "Teenage Fertility and High School Completion." *Review of Economics and Statistics* 76(3): 413–424.

[102] Ribar, David C. 1999. "The Socioeconomic Consequences of Young Women's Childbearing: Reconciling Disparate Evidence." *Journal of Population Economics* 12(4): 547–565.

[103] Roby, Douglas E. 2004. "Research on school attendance and student achievement: A study of Ohio schools." *Educational Research Quarterly 28.1*

[104] Rumberger, Russell. 2004. "Why Students Drop Out of School." In *Dropouts in America*, edited by Gary Orfield, pp. 243 54. Cambridge, Mass.: Harvard Education Press.

[105] Sanders, Seth, Jeffrey Smith and Ye Zhang. 2008. "Teenage Childbearing and Maternal Schooling Outcomes: Evidence from Matching." University of Michigan Working Paper.

[106] Smedley, Brian D., Adrienne Y. Stith, and Alan R. Nelson. 2003. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: National Academies Press.

[107] Smith, J., 2009. “Can regression discontinuity help answer an age-old question in education?The effect of age on elementary and secondary school achievement.” *The BE Journal of Economic Analysis & Policy, 9(1).*

[108] Staff, J., Schulenberg, J.E. and Bachman, J.G., 2010. “Adolescent work intensity, school performance, and academic engagement.” *Sociology of Education, 83(3),* pp.183-200.

[109] Stratton, L. S., OToole, D. M., & Wetzel, J. N. 2008. “A multinomial logit model of college stopout and dropout behavior.” *Economics of education review, 27(3),* 319-331.

[110] Todd, P. E., & Wolpin, K. I. 2003. “On the specification and estimation of the production function for cognitive achievement.” *The Economic Journal*, 113(485).

[111] Todd, Petra E. and Kenneth I. Wolpin. 2007. “The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps.” *Journal of Human Capital* 1(1): 91–136.

[112] Utah Education Policy Center. 2012. “Chronic Absenteeism.” *Research Brief, University of Utah* web: ¡http://www.utahdataalliance.org/downloads/ChronicAbsenteeismResearchBrief.pdf¿

[113] Wahistrom, K. 2002. “Changing times: Findings from the first longitudinal study of later high school start times.” *Nassp Bulletin*, 86(633), 3-21.

[114] Walker, Sarah Cusworth, Suzanne E.U. Kerns, Aaron R. Lyon, Eric J. Bruns, and T.J. Cosgrove. 2010. “Impact of School-Based Health Center Use on Academic Outcomes.” *Journal of Adolescent Health* 46: 251-257.

[115] Wong, Mitchell, and others. 2002. “Contribution of Major Diseases to Disparities in Mortality.” *New England Journal of Medicine* 347, no. 20: 1585 92.

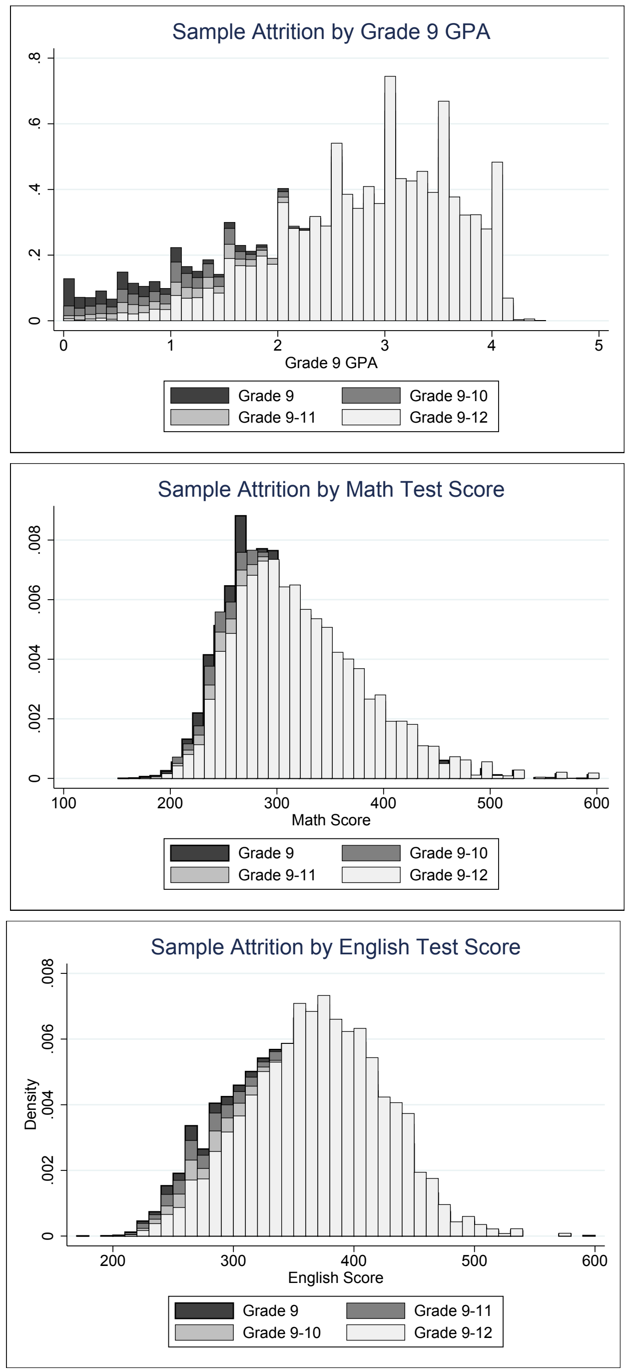
[116] Yakusheva, Olga and Jason Fletcher. 2015. “Learning from Teen Childbearing Experiences of Close Friends: Evidence using Miscarriages as a Natural Experiment.” *Review of Economics and Statistics* 97(1): 29-43.

APPENDIX A

APPENDIX FOR CHAPTER 1

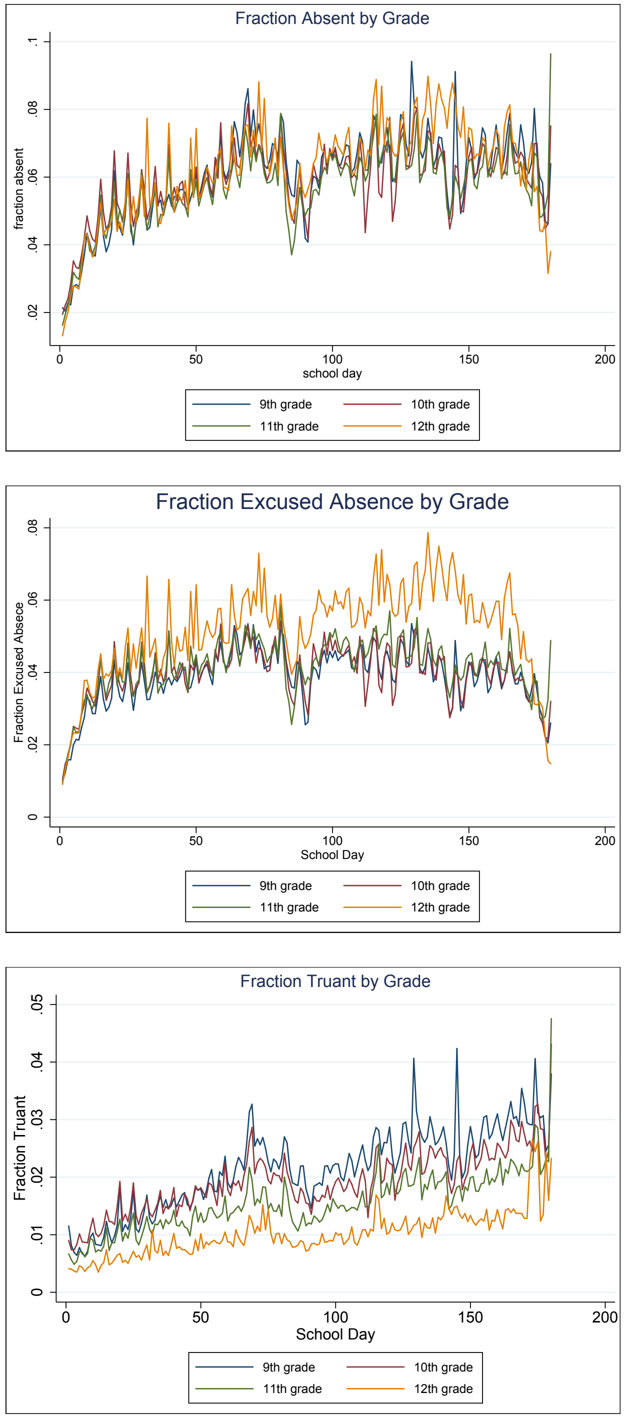
A.1 Additional Figures

Figure A.1: Attrition after 9th Grade



Histograms include all students who enter 9th grade and are color coded by the part of the 9th grade distribution that remains for grade 9 only, grades 9 and 10, grades 9,10, and 11, and all grades.

Figure A.2: School Year Absence Type Trends by Grade



Figures contain average daily attendance trends by grade for total absences (top), excused absences (middle), and truant absences(bottom). Each data point is a school day.

APPENDIX B

APPENDIX FOR CHAPTER 2

B.1 Additional Tables

Table B.1: Yearly Transition Results OLS: 9th Graders Continuing in Original School for 10th Grade

	(1) Continues	(2) Continues	(3) Continues	(4) Continues	(5) Continues
Total Truancies	-0.0089	-0.0089	-0.0093	-0.0056	-0.0062
	(0.000178)	(0.000185)	(0.000248)	(0.0002)	(0.000268)
ELA Score			0.0005		-.000058
			(0.00003)		(.000042)
GPA				0.0762	0.0658
				(0.0033)	(0.00359)
GPA under 2.0				-0.019	-0.0205
				(0.007)	(0.00662)
Demographics		X	X	X	X
Cohort FE	X	X	X	X	X
Observations	37229	33923	32280	32280	33923

Standard errors are clustered at the cohort (school by entrance year) level. Demographic variables include race, ethnicity, gender, ELL status, and high school entrance age. Students who transfer outside the district before the end of 9th grade are excluded from the sample.

Table B.2: Yearly Transition Results OLS: 9th Graders Continuing in Original School for 10th Grade with Excused Absences

	(1) Continues	(2) Continues	(3) Continues	(4) Continues	(5) Continues
Total Truancies	-0.0083	-0.0082	-0.0082	-0.0056	-0.0058
	(0.000178)	(0.000185)	(0.000249)	(0.0002)	(0.000267)
Total Absences	-0.0063	-0.0063	-0.0062	-0.0044	-0.0050
	(0.000240)	(0.000240)	(0.000246)	(0.00024)	(0.000249)
ELA Score			0.00043		-.000028
			(0.00003)		(.000042)
GPA				0.0658	0.0535
				(0.0034)	(0.0067)
GPA under 2.0				-0.0215	-0.0240
				(0.0097)	(0.00658)
Demographics		X	X	X	X
Cohort FE	X	X	X	X	X
Observations	37229	33923	32280	32280	33923

Standard errors are clustered at the cohort (school by entrance year) level. Demographic variables include race, ethnicity, gender, ELL status, and high school entrance age. Students who transfer outside the district before the end of 9th grade are excluded from the sample. Total absences are the total excused absences that a student has in addition to truancies.

Table B.3: Yearly Transition Results OLS: 11th Graders Continuing in Original School for 12th Grade

	(1) Continues	(2) Continues	(3) Continues	(4) Continues	(5) Continues
Total Truancies	-0.0104	-0.0104	-0.0129	-0.00620	-0.00753
	(0.0002)	(0.000216)	(0.00032)	(0.00217)	(0.00032)
ELA Score			0.00059		-.000028
			(0.00003)		(.000035)
GPA				0.0474	0.0434
				(0.0025)	(0.0029)
GPA under 2.0				-0.174	-0.169
				(0.006)	(0.00592)
Demographics		X	X	X	X
Cohort FE	X	X	X	X	X
Observations	25703	25661	24912	25661	24912

Standard errors are clustered at the cohort (school by entrance year) level. Demographic variables include race, ethnicity, gender, ELL status, and high school entrance age. Students who transfer outside the district before the end of 11th grade are excluded from the sample.

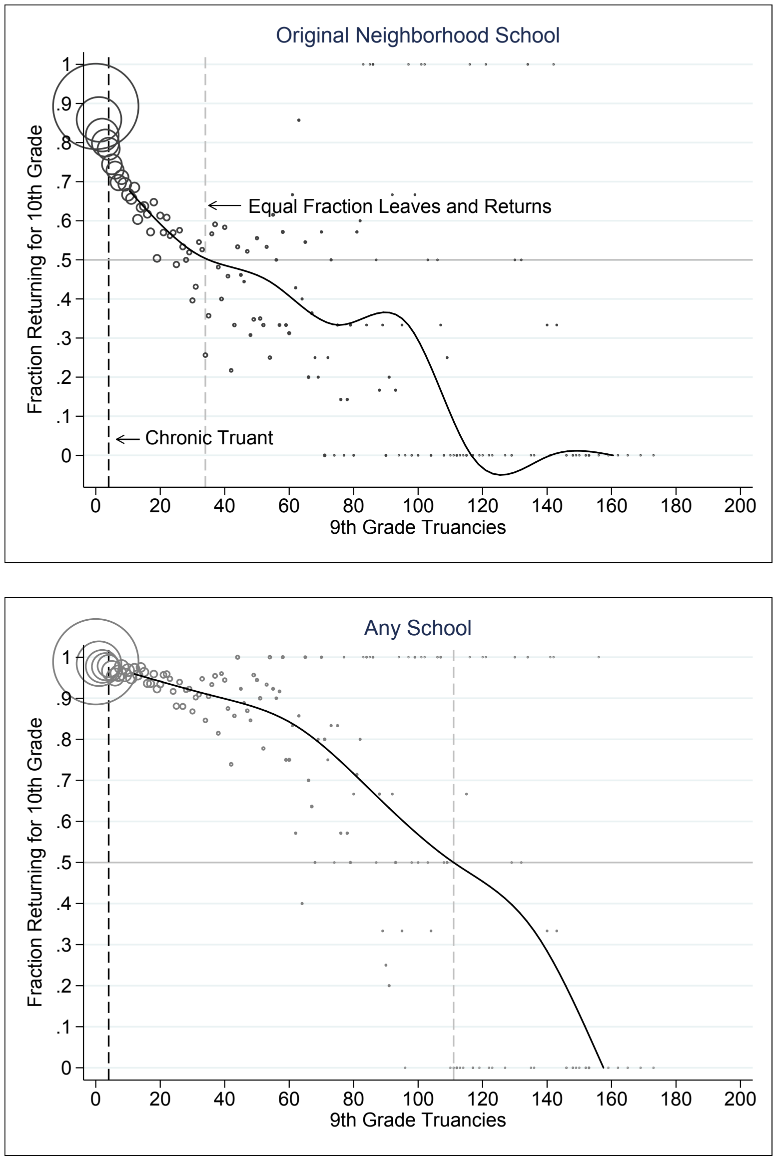
Table B.4: Yearly Transition Results OLS: 11th Graders Continuing in Original School for 12th Grade with Excused Absences

	(1) Continues	(2) Continues	(3) Continues	(4) Continues	(5) Continues
Total Truancies	-0.0103	-0.0101	-0.0125	-0.00614	-0.00742
	(0.0002)	(0.000216)	(0.00032)	(0.00217)	(0.00032)
Total Absences	-0.0026	-0.0030	-0.0020	-0.0014	-0.001
	(0.00023)	(0.000229)	(0.000227)	(0.00220)	(0.000219)
ELA Score			0.00057		-.000024
			(0.00003)		(.000035)
GPA				0.0452	0.0418
				(0.0026)	(0.0029)
GPA under 2.0				-0.175	-0.170
				(0.0061)	(0.00592)
Demographics		X	X	X	X
Cohort FE	X	X	X	X	X
Observations	25703	25661	24912	25661	24912

Standard errors are clustered at the cohort (school by entrance year) level. Demographic variables include race, ethnicity, gender, ELL status, and high school entrance age. Students who transfer outside the district before the end of 9th grade are excluded from the sample. Total absences are the total excused absences that a student has in addition to truancies.

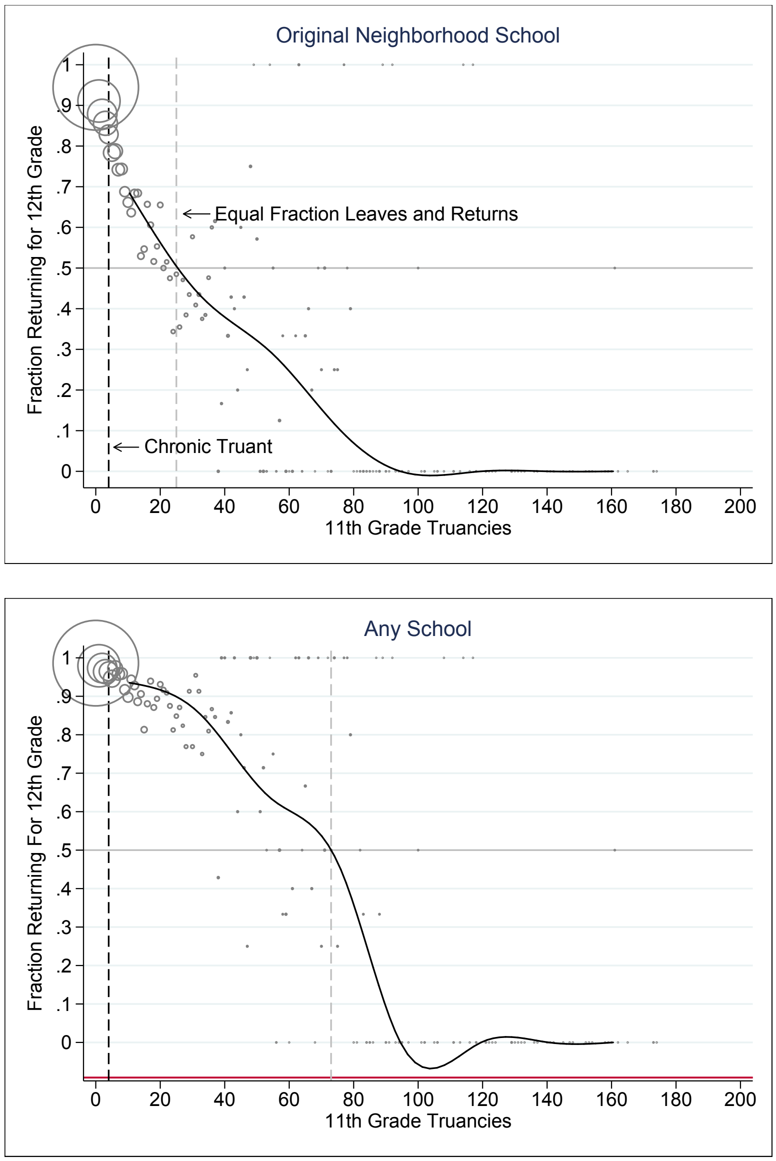
B.2 Additional Figures

Figure B.1: 9th to 10th Grade Transition



Each circle is weighted by the number of students in the sample with that number of truancies, the line is a median spline also weighted by number of students with each level of truancy. Vertical lines indicate the level at which students are considered chronic truants and the median spline estimate of the number of truancies when a student has an equal probability of returning or leaving her school. The top panel is the return to the original school, the bottom panel is the return to any school.

Figure B.2: 11th to 12th Grade Transition



Each circle is weighted by the number of students in the sample with that number of truancies, the line is a median spline also weighted by number of students with each level of truancy. Vertical lines indicate the level at which students are considered chronic truants and the median spline estimate of the number of truancies when a student has an equal probability of returning or leaving her school. The top panel is the return to the original school, the bottom panel is the return to any school.

APPENDIX C

APPENDIX FOR CHAPTER 3

C.1 Additional Tables

Table C.1: 1990 Census Characteristics of Counties and School Districts by Treatment Status

	Counties		School Districts	
Variable	Ever Treated	Never Treated	Ever Treated	Never Treated
Median Rent	466.49	429.43	401.09	370.35
Median Home Price	102146	79490	78517	69882
Median HH Income	31454	30946	26342	28363
Median Family Income	37057	36196	30950	32879
% Housing Occupied	90.64	91.30	87.70	87.23
% Urban	81.38	72.00	62.83	37.39
% Below Poverty	12.71	11.31	17.23	13.93
% w/ Public Assistance	7.39	6.07	9.67	7.13
% Inc. from Public Assistance	1.39	0.77	1.41	1.00
% w/ Wage Income	78.13	78.87	78.07	79.36
% Unemployed	4.10	3.60	5.18	4.05
% Not in Labor Force	34.07	33.86	37.34	36.93
% Male	48.58	48.91	48.78	49.21
% Black	11.79	8.85	11.93	4.71
% Hispanic	9.23	4.05	11.52	4.42
% Asian	2.55	1.36	1.81	0.80
% Other Race	1.06	0.63	1.90	1.74
% Under 6 Years Old	8.84	8.83	7.52	7.08
% 6-19 Years Old	19.54	20.26	22.18	22.55
% 20-34 Years Old	25.34	24.94	24.20	21.42
% 35-64 Years Old	33.96	33.88	33.28	34.65
% Institutionalized	1.32	1.32	1.24	1.07
% No HS Degree	22.90	22.31	28.55	25.98
% Some College	27.95	28.03	24.33	24.10
% BA+	19.81	17.84	15.31	14.42
% Married w/ Kids	25.94	28.50		
% Single w/ Kids	8.12	7.14		

Sources: 1990 Census Summary File 3 (counties) and School District Tabulation data. All tabulations use only the counties and school districts in the respective analysis sample.

Table C.2: The Effect of SBHC Services on Teen Birth and HS Dropout Rates Using Alternative Service Measures

Panel A: Births per 1,000 Aged 15-18	
Treatment Measure	(i)
Days Open per Week	-1.256*
	(0.742)
Hours Open per Week	-0.529**
	(0.167)

Panel B: High School Dropout Rate			
Treatment Measure	10^{th} Grade (i)	11^{th} Grade (ii)	12^{th} Grade (iii)
Days Open per Week	-0.038	-0.219*	-0.142
	(0.084)	(0.127)	(0.103)
Hours Open per Week	-0.003	-0.033	-0.011
	(0.016)	(0.020)	(0.018)
Average Days per Week		0.553	
Average Hours per Week		3.596	

Notes: Authors' estimates of equations (1) and (4) as described in the text. Each cell comes from a separate regression. All estimates in Panel A include county and state-by-year fixed effects and all estimates in Panel B include school district and state-by-year fixed effects. The 10^{th} grade dropout rate is calculated as 1 minus the ratio of diplomas awarded in year t and the 10^{th} grade enrollment in year $t-2$. The 11^{th} grade dropout rate equals 1 minus the ratio of diplomas awarded in year t and the 11^{th} grade enrollment in year $t-1$, and the 12^{th} grade dropout rate is calculated as 1 minus the ratio of diplomas awarded in year t and the 12^{th} grade enrollment in year t. Regressions are weighted by the high school aged population in the county (Panel A) or school district (Panel B). Standard errors clustered at the county (Panel A) or school district (Panel B) level are in parentheses: ** indicates significance at the 5% level and * indicates significance at the 10% level.

Table C.3: The Effect of SBHC Services on High School Dropout Rates (in Percent) – Diploma Data, by Birth Control Services

	Primary Care Staff Hours			Medical Staff Hours		
Grade:	10^{th}	11^{th}	12^{th}	10^{th}	11^{th}	12^{th}
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
<u>Birth Control Services</u>						
Hormones Prescribed On Site	-0.014	-0.004	0.001	-0.008	-0.002	0.001
	(0.012)	(0.014)	(0.014)	(0.006)	(0.007)	(0.007)
Hormones Referred, No Condoms	-0.0001	-0.050	-0.015	-0.001	-0.019	-0.008
	(0.001)	(0.041)	(0.025)	(0.010)	(0.016)	(0.011)
Hormones Referred & Condoms Dispensed	0.079	0.089	0.034	0.027	-0.029	-0.043
	(0.063)	(0.064)	(0.057)	(0.021)	(0.064)	(0.047)
No Birth Control Services	-0.001	-0.024	0.018	-0.014	-0.014	0.007
	(0.031)	(0.032)	(0.031)	(0.017)	(0.017)	(0.015)

Notes: Authors' estimates of equation (4) as described in the text. Each column comes from a separate regression. The birth control service measures include the number of service hours of each type in centers with the given birth control policy. The birth control policy groups are exhaustive and mutually exclusive. The 10^{th} grade dropout rate is calculated as 1 minus the ratio of diplomas awarded in year t and the 10^{th} grade enrollment in year $t-2$. The 11^{th} grade dropout rate equals 1 minus the ratio of diplomas awarded in year t and the 11^{th} grade enrollment in year $t-1$, and the 12^{th} grade dropout rate is calculated as 1 minus the ratio of diplomas awarded in year t and the 12^{th} grade enrollment in year t. All estimates include school district and state-by-year fixed effects, and the regressions are weighted by the high school aged population in the school district. Standard errors clustered at the school district level are in parentheses: ** indicates significance at the 5% level and * indicates significance at the 10% level.

Table C.4: The Effect of SBHC Services on High School Dropout Rates (in Percent) – Diploma Data Using Large Counties

Grade:	10^{th}	11^{th}	12^{th}
	(i)	(ii)	(iii)
Treatment Measure			
Center Indicator	0.546	0.153	0.642
	(0.492)	(0.483)	(0.392)
Primary Care Staff Hours	-0.004	-0.032	-0.010
	(0.011)	(0.021)	(0.013)
Medical Staff Hours	-0.004	-0.012*	-0.005
	(0.005)	(0.006)	(0.006)

Notes: Authors' estimates of equation (4) using NCES CCD high school diploma data from 1998-2010. The sample is comprised of the large counties that constitute the birth rate analysis sample. Each cell comes from a separate regression. The 10^{th} grade dropout rate is calculated as 1 minus the ratio of diplomas awarded in year t and the 10^{th} grade enrollment in year $t-2$. The 11^{th} grade dropout rate equals 1 minus the ratio of diplomas awarded in year t and the 11^{th} grade enrollment in year $t-1$, and the 12^{th} grade dropout rate is calculated as 1 minus the ratio of diplomas awarded in year t and the 12^{th} grade enrollment in year t. All estimates include school district and state-by-year fixed effects, and the regressions are weighted by the high school aged population in the school district. Standard errors clustered at the school district level are in parentheses: ** indicates significance at the 5% level and * indicates significance at the 10% level.

Table C.5: The Effect of SBHC Services on STD Rates per 1000 15-19 Year Olds

Panel A: Baseline Estimates			
	STDs	Chlamydia	Gonorrhea
Treatment Measure	(i)	(ii)	(iii)
Primary Care Staff Hours	-0.187	-0.102	-0.102
	(0.135)	(0.111)	(0.046)
Medical Staff Hours	-0.048	-0.021	-0.080
	(0.044)	(0.039)	(0.067)

Panel B: Controlling for Chlamydia and Gonorrhea Rates Among 25-29 Year Olds			
	STDs	Chlamydia	Gonorrhea
Treatment Measure	(i)	(ii)	(iii)
Primary Care Staff Hours	-0.187	-0.105	-0.077**
	(0.179)	(0.147)	(0.034)
Medical Staff Hours	-0.047	-0.019	-0.026**
	(0.052)	(0.043)	(0.010)

Notes: Authors' estimates of a version of equation (1) aggregated to the state-year level. Each cell comes from a separate regression. All estimates include state and year fixed effects. STD data are for years 1998-2011 and include chlamydia, gonorrhea and syphilis in column (i). Standard errors clustered at the state level are in parentheses: ** indicates significance at the 5% level and * indicates significance at the 10% level.

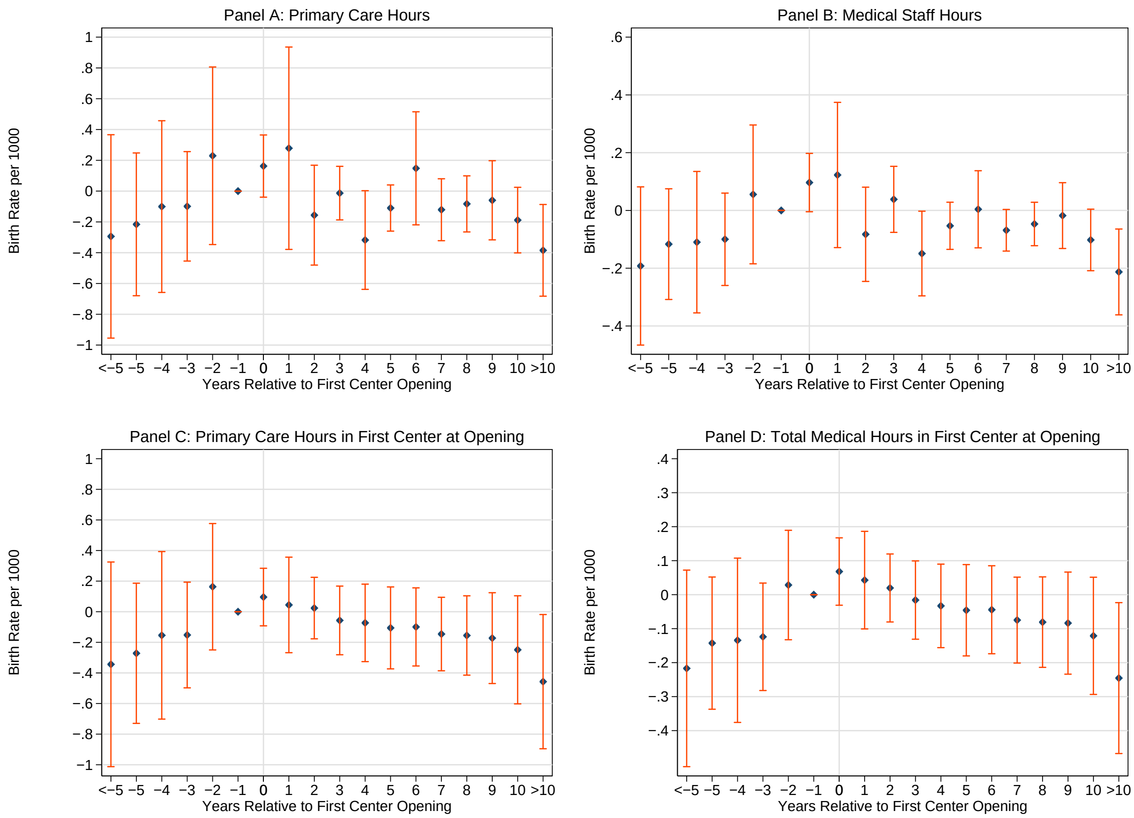
Table C.6: First-Stage Estimates from Instrumental Variables Models

	Birth Rates		Dropout Rates	
Service	Primary	Medical	Primary	Medical
Measure:	Care	Staff	Care	Staff
Center Indicator	2.657**	5.843**	4.919**	10.684**
	(0.221)	(0.474)	(0.441)	(0.933)
Time Since First Entry	0.033	-0.026	0.253**	0.289**
	(0.034)	(0.083)	(0.066)	(0.120)
(Time Since First Entry)^2	-0.0003	0.004	-0.002	-0.001
	(0.002)	(0.005)	(0.002)	(0.003)

Notes: Authors' estimates of equations (1) and (4) as described in the text. All estimates include county/school district and state-by-year fixed effects, and the regressions are weighted by the high school aged population in the county/school district. Standard errors clustered at the county/school district level are in parentheses: ** indicates significance at the 5% level and * indicates significance at the 10% level.

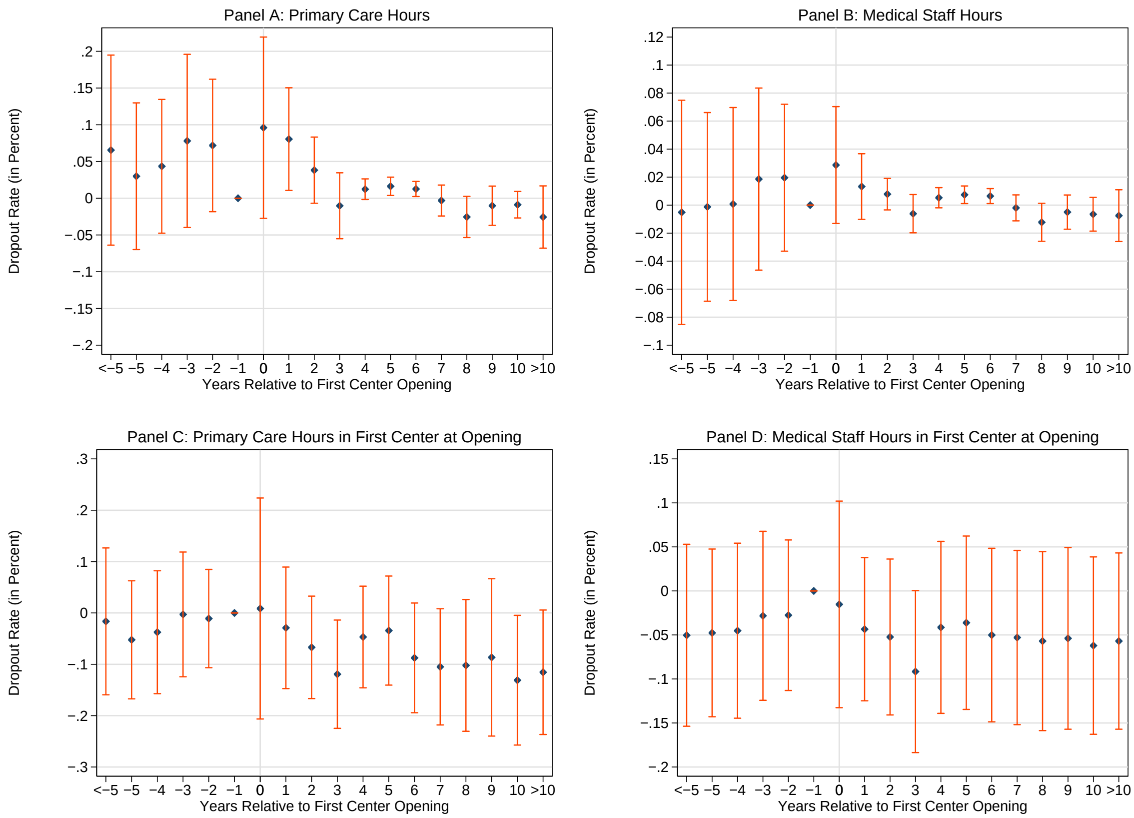
C.2 Additional Figures

Figure C.1: Event Study Estimates of the Effect of SBHC Services on Teen Birth Rates (per 1000 women)



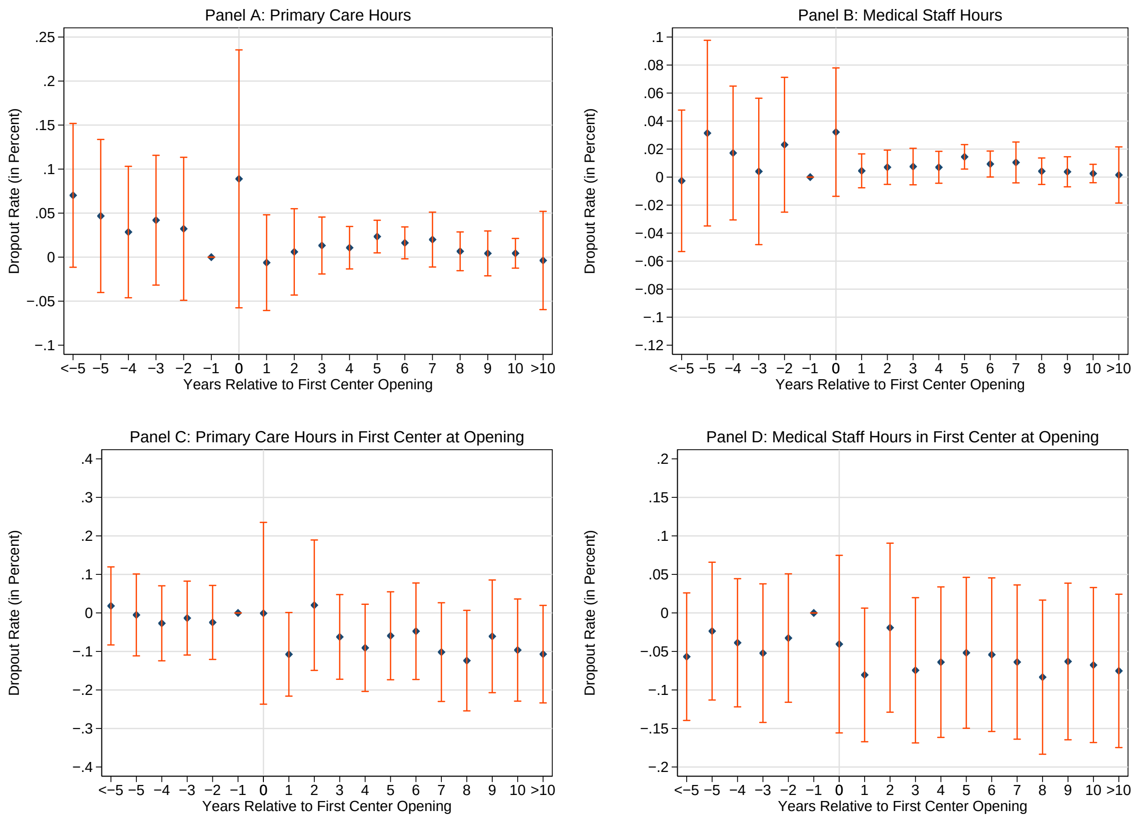
Authors' estimates of equation (3) as described in the text. The dependent variable in each panel is 15-18 year old birth rates per 1000. Each point shows the coefficient estimate on the service measure interacted with the relative time to the first center opening in the county. All estimates include county and state-by-year fixed effects, and the regressions are weighted by the high school aged population in the county. The lines extending from each point show the bounds of the 95% confidence intervals that are calculated using standard errors clustered at the county level. Relative year -1 is omitted, so all estimates are relative to this year.

Figure C.2: Event Study Estimates of the Effect of SBHC Primary Care and Medical Staff Services on 10^{th} Grade High School Dropout Rates (in Percent) – Diploma Data



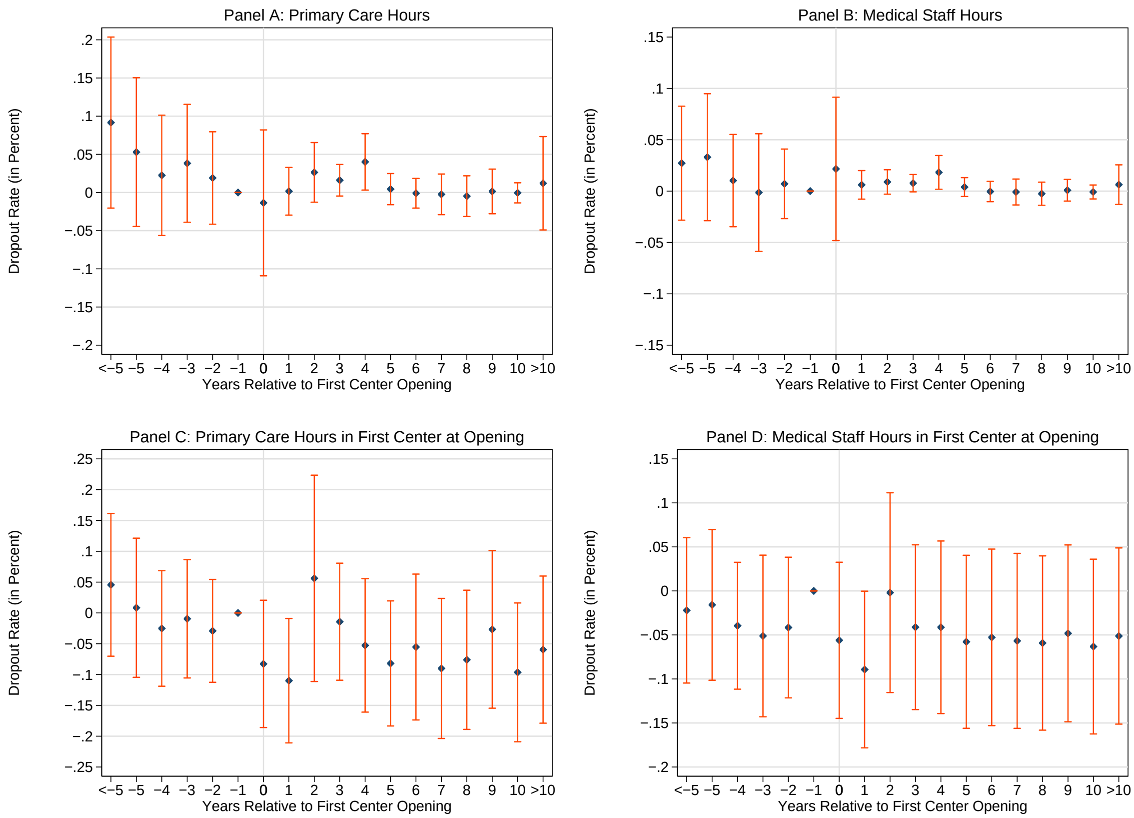
Authors' estimates of equation (6) as described in the text. Each point shows the coefficient estimate on the service hours measure interacted with the relative time to the first center opening in the school district. All estimates include school district and state-by-year fixed effects, and the regressions are weighted by the high school aged population in the school district. The lines extending from each point show the bounds of the 95% confidence intervals that are calculated using standard errors clustered at the school district level. Relative year -1 is omitted, so all estimates are relative to this year.

Figure C.3: Event Study Estimates of the Effect of SBHC Primary Care and Medical Staff Services on 11^{th} Grade High School Dropout Rates (in Percent) – Diploma Data



Authors' estimates of equation (6) as described in the text. Each point shows the coefficient estimate on the service hours measure interacted with the relative time to the first center opening in the school district. All estimates include school district and state-by-year fixed effects, and the regressions are weighted by the high school aged population in the school district. The lines extending from each point show the bounds of the 95% confidence intervals that are calculated using standard errors clustered at the school district level. Relative year -1 is omitted, so all estimates are relative to this year.

Figure C.4: Event Study Estimates of the Effect of SBHC Primary Care and Medical Staff Services on 12^{th} Grade High School Dropout Rates (in Percent) – Diploma Data



Authors' estimates of equation (6) as described in the text. Each point shows the coefficient estimate on the service hours measure interacted with the relative time to the first center opening in the school district. All estimates include school district and state-by-year fixed effects, and the regressions are weighted by the high school aged population in the school district. The lines extending from each point show the bounds of the 95% confidence intervals that are calculated using standard errors clustered at the school district level. Relative year -1 is omitted, so all estimates are relative to this year.